

# SUB-FINSLER GEOMETRY IN DIMENSION THREE

JEANNE N. CLELLAND AND CHRISTOPHER G. MOSELEY

**ABSTRACT.** We define the notion of *sub-Finsler geometry* as a natural generalization of sub-Riemannian geometry with applications to optimal control theory. We compute a complete set of local invariants, geodesic equations, and the Jacobi operator for the three-dimensional case and investigate homogeneous examples.

## 1. INTRODUCTION

Much attention has been given in recent years to sub-Riemannian geometry; it is a rich subject with many applications. In this paper we introduce the notion of *sub-Finsler geometry*, a natural generalization of sub-Riemannian geometry.

The motivation for this generalization comes from optimal control theory. A control system is usually presented in local coordinates as an underdetermined system of ordinary differential equations

$$(1.1) \quad \dot{x} = f(x, u),$$

where  $x \in \mathbb{R}^n$  represents the *state* of the system and  $u \in \mathbb{R}^s$  represents the *controls*, i.e., variables which may be specified freely in order to “steer” the system in a desired direction. More generally,  $x$  and  $u$  may take values in an  $n$ -dimensional manifold  $\mathcal{X}$  and an  $s$ -dimensional manifold  $\mathcal{U}$ , respectively. Typically there are constraints on how the system may be “steered” from one state to another, so that  $s < n$ . The systems of greatest interest are *controllable*, i.e., given any two states  $x_1, x_2$ , there exists a solution curve of (1.1) connecting  $x_1$  to  $x_2$ .

Consider the large class of systems which are linear (but not affine linear) in the control variables  $u$  and depend smoothly on the state variables  $x$ , i.e., systems of the form

$$(1.2) \quad \dot{x} = f(x)u,$$

where  $f(x)$  is a matrix whose entries are arbitrary smooth functions of  $x$ . This class is by no means all-inclusive, but it does contain many systems of interest; an example is given below. For such a system, *admissible* paths in the state space are those for which the tangent vector to the path at each point  $x \in \mathcal{X}$  is contained in the subspace  $D_x \subset T_x \mathcal{X}$  determined by the image of the  $n \times s$  matrix  $f(x)$ . Often this matrix is smooth and has constant rank  $s$ , in which case  $D$  is a rank  $s$  distribution on  $\mathcal{X}$ . (In this case the variables  $(x, u)$  may be regarded as local coordinates on the distribution  $(\mathcal{X}, D)$ .) Thus the admissible paths in the state space are precisely the *horizontal curves* of the distribution  $D$ , i.e., curves whose tangent vectors at each point are contained in  $D$ . By a theorem of Chow [7], the system (1.2) is controllable if and only if the distribution  $D$  on  $\mathcal{X}$  is *bracket-generating*, i.e., if the iterated brackets of vector fields contained in  $D$  span the entire tangent space at each point  $x \in \mathcal{X}$ .

---

2000 *Mathematics Subject Classification.* Primary(53C17, 53B40, 49J15), Secondary(58A15, 53C10).

*Key words and phrases.* sub-Finsler geometry, optimal control theory, exterior differential systems, Cartan’s method of equivalence.

Given a distribution  $(\mathcal{X}, D)$  representing a system of the form (1.2), we next consider the problem of *optimal control*: what is the most efficient path between two given points in  $\mathcal{X}$ ? In order to answer this question, we must have some measure of the cost required to move in the state space. This measure is typically specified by a first-order Lagrangian functional  $L$  defined on the horizontal curves of  $D$ : given a horizontal curve  $\gamma : [a, b] \rightarrow \mathcal{X}$ , the *action*  $\mathcal{L}(\gamma)$  is defined to be

$$\mathcal{L}(\gamma) = \int_{\gamma} L(x, \dot{x}) dx = \int_{\gamma} \bar{L}(x, u) dx,$$

where, since  $\gamma$  is a solution curve of (1.2), we define  $\bar{L}(x, u) = L(x, f(x)u)$ . Often the Lagrangian has the form

$$\bar{L}(x, u) = \sqrt{g_{ij}(x)u^i u^j}$$

(summation on repeated indices being understood), and in this case it defines a *sub-Riemannian metric*  $\langle \cdot, \cdot \rangle$  on  $D$  (i.e., a Riemannian metric on each subspace  $D_x \subset T_x \mathcal{X}$ ) in the obvious way. Horizontal paths which minimize the action functional are precisely the geodesics of the sub-Riemannian metric.

**Example 1.1.** Consider a wheel rolling without slipping on the Euclidean plane  $\mathbb{E}^2$ . The wheel's configuration can be represented by the vector  ${}^t(x, y, \varphi, \psi)$ , where  $(x, y)$  is the wheel's point of contact with the plane,  $\phi$  is the angle of rotation of a marked point on the wheel from the vertical, and  $\psi$  is the wheel's heading angle, i.e., the angle made by the tangent line to the curve traced by the wheel on the plane with the  $x$ -axis. Thus the state space has dimension four and is naturally isomorphic to  $\mathbb{R}^2 \times S^1 \times S^1$ .

The condition that the wheel rolls without slipping is equivalent to the statement that its path  ${}^t(x(t), y(t), \varphi(t), \psi(t))$  in the state space satisfies the differential equation

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\varphi} \\ \dot{\psi} \end{bmatrix} = u_1(t) \begin{bmatrix} \cos \psi \\ \sin \psi \\ 1 \\ 0 \end{bmatrix} + u_2(t) \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

for some control functions  $u_1(t), u_2(t)$ . Thus the velocity vector  ${}^t(\dot{x}, \dot{y}, \dot{\varphi}, \dot{\psi})$  of any solution curve must lie in the distribution  $D$  spanned by the vector fields

$$V_1 = (\cos \psi) \frac{\partial}{\partial x} + (\sin \psi) \frac{\partial}{\partial y} + \frac{\partial}{\partial \varphi}$$

$$V_2 = \frac{\partial}{\partial \psi}.$$

A natural sub-Riemannian metric on  $D$  is obtained by declaring the vector fields  $V_1, V_2$  to be orthonormal, i.e., by setting

$$\langle u_1 V_1 + u_2 V_2, u_1 V_1 + u_2 V_2 \rangle = u_1^2 + u_2^2.$$

The integral of this quadratic form measures the work done in rotating the heading angle  $\psi$  at the rate  $\dot{\psi}$  and propelling the wheel forward at the rate  $\dot{\varphi}$ .

But what if the natural measure on horizontal curves is not the square root of a quadratic form? For instance, suppose we modified Example 1.1 by rolling the wheel on an *inclined* plane? (Assume that the wheel has sufficient friction to remain motionless if no energy is put into the system.) We would expect more energy to be required to move the wheel

uphill than downhill, so the natural Lagrangian would not even be symmetric in  $u$  (i.e., it would not satisfy the condition  $\bar{L}(x, -u) = \bar{L}(x, u)$ ), let alone be the square root of a quadratic form in  $u$ . It is not difficult to imagine examples where the dependence of  $\bar{L}$  on  $u$  becomes quite complicated as  $u$  changes direction. This leads us to generalize the notion of a sub-Riemannian metric on  $(\mathcal{X}, D)$  by replacing the Riemannian metric on each subspace  $D_x \subset T_x\mathcal{X}$  with a Finsler metric.

Recall that a *Finsler metric* on a manifold  $\mathcal{M}$  is a function

$$F : T\mathcal{M} \rightarrow [0, \infty)$$

with the following properties:

- (1) Regularity:  $F$  is  $C^\infty$  on the slit tangent bundle  $T\mathcal{M} \setminus 0$ .
- (2) Positive homogeneity:  $F(x, \lambda y) = \lambda F(x, y)$  for all  $\lambda > 0$ . (Here  $x$  is any system of local coordinates on  $\mathcal{M}$  and  $(x, y)$  is the corresponding canonical coordinate system on  $T\mathcal{M}$ .)
- (3) Strong convexity: The  $n \times n$  Hessian matrix

$$\left[ \frac{\partial^2 (\frac{1}{2}F^2)}{\partial y^i \partial y^j} \right]$$

is positive definite at every point of  $T\mathcal{M} \setminus 0$ .

(For details, see [1].) In other words, a Finsler metric on a manifold  $\mathcal{M}$  is a smoothly varying Minkowski norm on each tangent space  $T_x\mathcal{M}$ .

Condition 3 implies that the “unit sphere” in each tangent space  $T_x\mathcal{M}$  (also known as the *indicatrix* for the Finsler metric on  $T_x\mathcal{M}$ ) is a smooth, strictly convex hypersurface enclosing the origin  $0_x \in T_x\mathcal{M}$ . The converse is almost – but not quite – true: there exist strictly convex hypersurfaces for which the corresponding Hessian matrix is only positive semi-definite along a closed subset; see [1] for examples. We will say that a hypersurface  $\Sigma_x \subset T_x\mathcal{M}$  which encloses the origin is *strongly convex* if it is the indicatrix for a Minkowski norm on  $T_x\mathcal{M}$ ; thus strong convexity implies strict convexity, but not vice-versa.

In the Riemannian case, the indicatrix must be an ellipsoid centered at  $0_x$ , but in the Finsler case it may be much more general. In particular, it need not be symmetric about the origin.

We are now ready to define our primary object of study.

**Definition 1.2.** A sub-Finsler metric on a smooth distribution  $D$  of rank  $s$  on an  $n$ -dimensional manifold  $\mathcal{X}$  is a smoothly varying Finsler metric on each subspace  $D_x \subset T_x\mathcal{X}$ . A sub-Finsler manifold, denoted by the triple  $(\mathcal{X}, D, F)$ , is a smooth  $n$ -dimensional manifold  $\mathcal{X}$  equipped with a sub-Finsler metric  $F$  on a bracket-generating distribution  $D$  of rank  $s > 0$ . The length of a horizontal curve  $\gamma : [a, b] \rightarrow \mathcal{X}$  is

$$\mathcal{L}(\gamma) = \int_a^b F(\dot{\gamma}(t)) dt.$$

Replacing the Riemannian metric on  $D$  by a Finsler metric allows more general action functionals to be considered. The rather stringent requirement that the Lagrangian be the square root of a quadratic form is replaced by the more natural requirement that it be positive-homogeneous in  $u$  (which is necessary if the length of an oriented curve is to be independent of parametrization), and that it be strongly convex (which is necessary if there

are to exist locally minimizing paths in every direction). The problem of finding minimizing paths satisfying (1.2) is equivalent to finding geodesics of the sub-Finsler manifold  $(\mathcal{X}, D, F)$ .

In this paper we will investigate sub-Finsler manifolds in the simplest nontrivial case: a three-dimensional manifold  $\mathcal{X}$  with a rank two contact distribution  $D$ . We will work locally, and thus we will not generally concern ourselves with the issue of local vs. global existence of objects such as coordinates, vector fields, and differential forms.

In the next two sections we will review some results of Huguen [11] concerning sub-Riemannian geometry in dimension three and some results of Cartan [3, 6] concerning the geometry of Finsler surfaces. We will then combine these techniques to construct a complete set of local invariants for sub-Finsler manifolds in dimension three via Élie Cartan's method of equivalence. (See [8] for an exposition of this method. The reader should be aware that where Gardner uses left group actions, we use right group actions for greater ease of computation.) Additionally, we will derive the geodesic equations, compute the Jacobi operator for the second variation problem, and investigate homogeneous examples.

## 2. REVIEW OF SUB-RIEMANNIAN GEOMETRY OF 3-MANIFOLDS

The material in this section is taken from Keener Huguen's Ph.D. thesis [11]. Unfortunately this thesis was never published, but some of the results are summarized in [14].

Let  $(\mathcal{X}, D, \langle, \rangle)$  be a sub-Riemannian structure on a 3-manifold  $\mathcal{X}$  with a contact distribution  $D$ . A local coframing  $(\eta^1, \eta^2, \eta^3)$  on  $\mathcal{X}$  is said to be *0-adapted* to the sub-Riemannian structure if  $D = \{\eta^3\}^\perp$  and  $\langle, \rangle = (\eta^1)^2 + (\eta^2)^2$ . The set of 0-adapted coframings of  $\mathcal{X}$  forms a  $G_0$ -structure  $\mathcal{B}_0 \rightarrow \mathcal{X}$ , where  $G_0$  is the Lie group

$$G_0 = \left\{ \begin{bmatrix} A & b \\ 0 & c \end{bmatrix} : A \in O(2), b \in \mathbb{R}^2, c \in \mathbb{R}^* \right\}.$$

We apply the method of equivalence to this  $G_0$ -structure, and after two reductions we arrive at the bundle of *2-adapted* coframings. This is a  $G_2$ -structure  $\mathcal{B}_2 \rightarrow \mathcal{X}$ , where  $G_2$  is the Lie group

$$G_2 = \left\{ \begin{bmatrix} A & 0 \\ 0 & \det A \end{bmatrix} : A \in O(2) \right\}.$$

There is a canonical coframing  $(\omega^1, \omega^2, \omega^3, \alpha)$  (also known as an *(e)-structure*) on  $\mathcal{B}_2$  whose structure equations are

$$\begin{aligned} d\omega^1 &= -\alpha \wedge \omega^2 + A_1 \omega^2 \wedge \omega^3 + A_2 \omega^3 \wedge \omega^1 \\ d\omega^2 &= \alpha \wedge \omega^1 + A_2 \omega^2 \wedge \omega^3 - A_1 \omega^3 \wedge \omega^1 \\ d\omega^3 &= \omega^1 \wedge \omega^2 \\ d\alpha &= S_1 \omega^2 \wedge \omega^3 + S_2 \omega^3 \wedge \omega^1 + K \omega^1 \wedge \omega^2. \end{aligned} \tag{2.1}$$

Differentiating these equations shows that

$$\begin{aligned} dA_1 &= -2A_2\alpha + \sum_{i=1}^3 B_{1i} \omega^i \\ dA_2 &= 2A_1\alpha + \sum_{i=1}^3 B_{2i} \omega^i \end{aligned}$$

for some functions  $B_{ij}$  on  $\mathcal{B}_2$ , and that

$$S_1 = B_{12} - B_{21}, \quad S_2 = B_{11} + B_{22}.$$

By the general theory of  $(e)$ -structures, the functions  $A_1, A_2, K$  form a complete set of differential invariants for the  $G_2$ -structure  $\mathcal{B}_2 \rightarrow \mathcal{X}$ , and hence for the sub-Riemannian structure  $(\mathcal{X}, D, \langle, \rangle)$ .

For later use, we observe that  $\mathcal{B}_2$  may be viewed geometrically as a double cover of the unit circle bundle of the sub-Riemannian metric. If the sub-Riemannian structure  $(\mathcal{X}, D, \langle, \rangle)$  is *orientable* (i.e., if we can choose an orientation on each of the subspaces  $D_x$  which varies smoothly on  $\mathcal{X}$ ), then  $\mathcal{B}_2$  consists of two disjoint connected components. In this case we can restrict the set of 0-adapted coframings by requiring that such a coframing be *oriented*, i.e., that the 2-form  $\eta^1 \wedge \eta^2$  be a positive area form on  $D$ . Doing so replaces the  $O(2)$  component of the structure group by  $SO(2)$ . This does not change anything essential in the preceding discussion, but it does lead to a  $G_2$ -structure  $\mathcal{B}_2$  which is connected and is naturally isomorphic to the unit circle bundle of  $(\mathcal{X}, D, \langle, \rangle)$ .

### 3. REVIEW OF FINSLER GEOMETRY OF SURFACES

The material in this section is taken from [3]. (We will, however, use the more standard notation for the invariants which is found in [1].)

A Finsler metric on a surface  $\mathcal{M}$  is determined by its indicatrix bundle: this is a smooth hypersurface  $\Sigma^3 \subset T\mathcal{M}$  with the property that each fiber  $\Sigma_x = \Sigma \cap T_x\mathcal{M}$  is a smooth, strongly convex curve which surrounds the origin  $0_x \in T_x\mathcal{M}$ . A 3-manifold  $\Sigma \subset T\mathcal{M}$  satisfying this condition is called a *Finsler structure* on  $\mathcal{M}$ . A differentiable curve  $\gamma : [a, b] \rightarrow \mathcal{M}$  is called a  $\Sigma$ -curve if, for every  $s \in [a, b]$ , the velocity vector  $\gamma'(s)$  lies in  $\Sigma$ . The following result is taken from [3] and is due to Cartan [6]:

**Proposition .** *Let  $\Sigma \subset T\mathcal{M}$  be a Finsler structure on an oriented surface  $\mathcal{M}$ , with basepoint projection  $\pi : \Sigma \rightarrow \mathcal{M}$ . Then there exists a unique coframing  $(\omega^1, \omega^2, \alpha)$  on  $\Sigma$  with the following properties:*

- (1)  $\omega^1 \wedge \omega^2$  is a positive multiple of any  $\pi$ -pullback of a positive 2-form on  $\mathcal{M}$ .
- (2) The tangential lift  $\gamma'$  of any  $\Sigma$ -curve satisfies  $(\gamma')^*\omega^2 = 0$  and  $(\gamma')^*\omega^1 = dt$ .
- (3)  $d\omega^1 \wedge \omega^2 = d\omega^2 \wedge \alpha = 0$ .
- (4)  $\omega^1 \wedge d\omega^1 = \omega^2 \wedge d\omega^2$ .
- (5)  $d\omega^1 = -\alpha \wedge \omega^2$ .

Moreover, there exist functions  $I, J, K$  on  $\Sigma$  such that

$$\begin{aligned} d\omega^1 &= -\alpha \wedge \omega^2 \\ d\omega^2 &= \alpha \wedge \omega^1 - I \alpha \wedge \omega^2 \\ d\alpha &= K \omega^1 \wedge \omega^2 + J \alpha \wedge \omega^2. \end{aligned} \tag{3.1}$$

The Finsler structure on  $\mathcal{M}$  is Riemannian if and only if  $I \equiv 0$ ; in this case, differentiating (3.1) shows that  $J \equiv 0$  as well, and we recover the familiar structure equations

$$\begin{aligned} d\omega^1 &= -\alpha \wedge \omega^2 \\ d\omega^2 &= \alpha \wedge \omega^1 \\ d\alpha &= K \omega^1 \wedge \omega^2 \end{aligned}$$

for an orthonormal coframing  $(\omega^1, \omega^2)$  on  $\mathcal{M}$ . In this case,  $\alpha$  is the Levi-Civita connection form, and  $K$  is the usual Gauss curvature on the surface. For general Finsler surfaces, the function  $K$  (called the *flag curvature*) is a well-defined function only on  $\Sigma$ , not on  $\mathcal{M}$ .

#### 4. THE SUB-FINSLER EQUIVALENCE PROBLEM

Let  $(\mathcal{X}, D, F)$  be a sub-Finsler manifold consisting of a three-dimensional manifold  $\mathcal{X}$ , a rank two contact distribution  $D$  on  $\mathcal{X}$ , and a sub-Finsler metric  $F$  on  $D$ . (Recall that  $D$  is *contact* if, for any two vector fields  $\mathbf{v}_1, \mathbf{v}_2$  locally spanning  $D$ , the vectors  $\mathbf{v}_1, \mathbf{v}_2$ , and  $[\mathbf{v}_1, \mathbf{v}_2]$  span the tangent space of  $\mathcal{X}$  at each point.) As in the Finsler case, the sub-Finsler metric  $F$  is completely determined by its indicatrix bundle

$$\Sigma = \{\mathbf{u} \in D \mid F(\mathbf{u}) = 1\}.$$

$\Sigma$  has dimension four, and each fiber  $\Sigma_x = \Sigma \cap D_x$  is a smooth, strongly convex curve in  $D_x$  which surrounds the origin  $0_x \in D_x$ . A 4-manifold  $\Sigma \subset T\mathcal{X}$  satisfying this condition will be called a *sub-Finsler structure* on  $(\mathcal{X}, D)$ .

We will compute invariants for sub-Finsler structures via Cartan's method of equivalence. We begin by constructing a coframing on  $\Sigma$  which is nicely adapted to the sub-Finsler structure; this procedure closely follows that used in [2] for constructing an adapted coframing for a Finsler structure on a surface.

Let  $g$  be any fixed sub-Riemannian metric on  $(\mathcal{X}, D)$ , and let  $\Sigma_1$  be the unit circle bundle for  $g$ . Then there exists a well-defined, smooth function  $r : \Sigma_1 \rightarrow \mathbb{R}^+$  with the property that

$$\Sigma = \{r(\mathbf{u})^{-1} \mathbf{u} \mid \mathbf{u} \in \Sigma_1\}.$$

Let  $\rho : \Sigma \rightarrow \Sigma_1$  be the diffeomorphism which is the inverse of the scaling map defined by  $r$ ; i.e.,  $\rho$  satisfies

$$\rho(r(\mathbf{u})^{-1} \mathbf{u}) = \mathbf{u}$$

for  $\mathbf{u} \in \Sigma_1$ .

Let  $\pi : \Sigma \rightarrow \mathcal{X}$ ,  $\pi_1 : \Sigma_1 \rightarrow \mathcal{X}$  denote the respective basepoint projections, and let  $\mathbf{u} \in \Sigma$ . (We trust that using the same notation for points in  $\Sigma$  and in  $\Sigma_1$  will not cause undue confusion.) We will say that a vector  $\mathbf{v} \in T_{\mathbf{u}}\Sigma$  is *monic* if  $\pi'(\mathbf{u})(\mathbf{v}) = \mathbf{u}$ . Since  $\pi'(\mathbf{u}) : T_{\mathbf{u}}\Sigma \rightarrow T_{\pi(\mathbf{u})}\mathcal{X}$  is surjective with a one-dimensional kernel, the set of monic vectors in  $T_{\mathbf{u}}\Sigma$  is an affine line. A nonvanishing 1-form  $\theta$  on  $\Sigma$  will be called *null* if  $\theta(\mathbf{v}) = 0$  for all monic vectors  $\mathbf{v}$ , and a 1-form  $\omega$  on  $\Sigma$  will be called *monic* if  $\omega(\mathbf{v}) = 1$  for all monic vectors  $\mathbf{v}$ . The set of null 1-forms spans a two-dimensional subspace of  $T_{\mathbf{u}}^*\Sigma$  at each point  $\mathbf{u} \in \Sigma$ , and the difference of any two monic 1-forms is a null form.

In the sub-Riemannian case,  $\omega^1$  is a monic form and the null 1-forms are spanned by  $\omega^2$  and  $\omega^3$ . (Recall that these forms are part of the canonical coframing on  $\Sigma_1$  described in section 2.) Moreover,  $D$  is defined by  $D = \{\omega^3\}^\perp$ ; this makes sense because according to the structure equations (2.1),  $\omega^3$  descends to a well-defined form on  $\mathcal{X}$ . Since the diagram

$$\begin{array}{ccc} \Sigma & \xrightarrow{\rho} & \Sigma_1 \\ \pi \searrow & & \swarrow \pi_1 \\ & \mathcal{X} & \end{array}$$

commutes, it is straightforward to verify that the null forms on  $\Sigma$  are spanned by  $\rho^*(\omega^2)$  and  $\rho^*(\omega^3)$ , that  $D = \{\rho^*(\omega^3)\}^\perp$ , and that  $\rho^*(r\omega^1)$  is a monic form on  $\Sigma$ .

A local coframing  $(\bar{\eta}^1, \bar{\eta}^2, \bar{\eta}^3, \bar{\phi})$  on  $\Sigma$  will be called *0-adapted* if it satisfies the conditions that  $\bar{\eta}^1$  is a monic form,  $\bar{\eta}^2$  and  $\bar{\eta}^3$  are null forms, and  $D = \{\bar{\eta}^3\}^\perp$ . For example, the coframing

$$(4.1) \quad \bar{\eta}^1 = \rho^*(r\omega^1), \quad \bar{\eta}^2 = \rho^*(\omega^2), \quad \bar{\eta}^3 = \rho^*(\omega^3), \quad \bar{\phi} = \rho^*(\alpha)$$

is 0-adapted. Any two 0-adapted coframings on  $\Sigma$  vary by a transformation of the form

$$(4.2) \quad \begin{bmatrix} \tilde{\eta}^1 \\ \tilde{\eta}^2 \\ \tilde{\eta}^3 \\ \tilde{\phi} \end{bmatrix} = \begin{bmatrix} 1 & a_1 & a_2 & 0 \\ 0 & b_1 & b_2 & 0 \\ 0 & 0 & b_3 & 0 \\ c_1 & c_2 & c_3 & c_4 \end{bmatrix}^{-1} \begin{bmatrix} \bar{\eta}^1 \\ \bar{\eta}^2 \\ \bar{\eta}^3 \\ \bar{\phi} \end{bmatrix}$$

with  $b_1 b_3 c_4 \neq 0$ . The set of all 0-adapted coframings forms a principal fiber bundle  $\mathcal{B}_0 \rightarrow \Sigma$ , with structure group  $G_0$  consisting of all matrices of the form (4.2). The right action of  $G_0$  on sections  $\sigma : \Sigma \rightarrow \mathcal{B}_0$  is given by  $g \cdot \sigma = g^{-1}\sigma$ . (This explains the inverse occurring in (4.2).)

There exist canonical 1-forms  $\eta^1, \eta^2, \eta^3, \phi$  on  $\mathcal{B}_0$  with the *reproducing property* that for any local section  $\sigma : \Sigma \rightarrow \mathcal{B}_0$ ,

$$\sigma^*(\eta^i) = \bar{\eta}^i, \quad \sigma^*(\phi) = \bar{\phi}.$$

These are referred to as the *semi-basic* forms on  $\mathcal{B}_0$ . A standard argument shows that there also exist (non-unique) 1-forms  $\alpha_i, \beta_i, \gamma_i$  (referred to as *pseudo-connection forms* or, more succinctly, *connection forms*), linearly independent from the semi-basic forms, and functions  $T_{jk}^i$  on  $\mathcal{B}_0$  (referred to as *torsion functions*) such that

$$(4.3) \quad \begin{bmatrix} d\eta^1 \\ d\eta^2 \\ d\eta^3 \\ d\phi \end{bmatrix} = - \begin{bmatrix} 0 & \alpha_1 & \alpha_2 & 0 \\ 0 & \beta_1 & \beta_2 & 0 \\ 0 & 0 & \beta_3 & 0 \\ \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \end{bmatrix} \wedge \begin{bmatrix} \eta^1 \\ \eta^2 \\ \eta^3 \\ \phi \end{bmatrix} + \begin{bmatrix} T_{10}^1 \eta^1 \wedge \phi \\ T_{10}^2 \eta^1 \wedge \phi \\ T_{12}^3 \eta^1 \wedge \eta^2 \\ 0 \end{bmatrix}.$$

These are the *structure equations* of the  $G_0$ -structure  $\mathcal{B}_0$ . The semi-basic forms and connection forms together form a local coframing on  $\mathcal{B}_0$ .

We proceed with the method of equivalence by examining how the functions  $T_{jk}^i$  vary if we change from one 0-adapted coframing to another. A straightforward computation shows that under a transformation of the form (4.2), we have

$$(4.4) \quad \begin{aligned} \tilde{T}_{10}^1 &= c_4 T_{10}^1 - \frac{a_1 c_4}{b_1} T_{10}^2 \\ \tilde{T}_{10}^2 &= \frac{c_4}{b_1} T_{10}^2 \\ \tilde{T}_{12}^3 &= \frac{b_1}{b_3} T_{12}^3. \end{aligned}$$

In particular, the functions  $T_{10}^2, T_{12}^3$  are *relative invariants*: if they vanish for any 0-adapted coframing, then they vanish for every 0-adapted coframing. The coframing (4.1) has  $T_{10}^2 = -r^{-1}$ ,  $T_{12}^3 = r^{-1}$ , so we can assume that these invariants are nonzero. (4.4) then implies that we can adapt coframings to arrange that

$$T_{10}^1 = 0, \quad T_{10}^2 = -1, \quad T_{12}^3 = 1.$$

A coframing satisfying this condition will be called *1-adapted*. For example, if we set

$$dr = r_1 \omega^1 + r_2 \omega^2 + r_3 \omega^3 + r_0 \phi,$$

then the coframing

$$(4.5) \quad \bar{\eta}^1 = \rho^*(r\omega^1 - r_0\omega^2), \quad \bar{\eta}^2 = \rho^*(r\omega^2), \quad \bar{\eta}^3 = \rho^*(r^2\omega^3), \quad \bar{\phi} = \rho^*(\alpha)$$

is 1-adapted. Any two 1-adapted coframings on  $\Sigma$  vary by a transformation of the form

$$(4.6) \quad \begin{bmatrix} \tilde{\eta}^1 \\ \tilde{\eta}^2 \\ \tilde{\eta}^3 \\ \tilde{\phi} \end{bmatrix} = \begin{bmatrix} 1 & 0 & a_2 & 0 \\ 0 & b_1 & b_2 & 0 \\ 0 & 0 & b_1 & 0 \\ c_1 & c_2 & c_3 & b_1 \end{bmatrix}^{-1} \begin{bmatrix} \bar{\eta}^1 \\ \bar{\eta}^2 \\ \bar{\eta}^3 \\ \bar{\phi} \end{bmatrix}$$

with  $b_1 \neq 0$ . The set of all 1-adapted coframings forms a principal fiber bundle  $\mathcal{B}_1 \subset \mathcal{B}_0$ , with structure group  $G_1$  consisting of all matrices of the form (4.6). When restricted to  $\mathcal{B}_1$ , the connection forms  $\alpha_1, \beta_3 - \beta_1, \gamma_4 - \beta_1$  become semi-basic, thereby introducing new torsion terms into the structure equations of  $\mathcal{B}_1$ . By adding multiples of the semi-basic forms to the connection forms so as to absorb as much of the torsion as possible, we can arrange that the structure equations of  $\mathcal{B}_1$  take the form

$$(4.7) \quad \begin{bmatrix} d\eta^1 \\ d\eta^2 \\ d\eta^3 \\ d\phi \end{bmatrix} = - \begin{bmatrix} 0 & 0 & \alpha_2 & 0 \\ 0 & \beta_1 & \beta_2 & 0 \\ 0 & 0 & \beta_1 & 0 \\ \gamma_1 & \gamma_2 & \gamma_3 & \beta_1 \end{bmatrix} \wedge \begin{bmatrix} \eta^1 \\ \eta^2 \\ \eta^3 \\ \phi \end{bmatrix} + \begin{bmatrix} T_{12}^1 \eta^1 \wedge \eta^2 + T_{20}^1 \eta^2 \wedge \phi \\ -\eta^1 \wedge \phi \\ \eta^1 \wedge \eta^2 + T_{13}^3 \eta^1 \wedge \eta^3 + T_{30}^3 \eta^3 \wedge \phi \\ 0 \end{bmatrix}.$$

Moreover, we have

$$\begin{aligned} 0 &\equiv d(d\eta^3) \mod \eta^3 \\ &\equiv T_{30}^3 \eta^1 \wedge \eta^2 \wedge \phi; \end{aligned}$$

therefore,  $T_{30}^3 = 0$ .

We now repeat this process. Under a transformation of the form (4.6), we have

$$(4.8) \quad \begin{aligned} \tilde{T}_{20}^1 &= b_1^2 T_{20}^1 \\ \tilde{T}_{12}^1 &= b_1 T_{12}^1 - b_1 c_1 T_{20}^1 - a_2 \\ \tilde{T}_{13}^3 &= T_{13}^3 + \frac{2b_2 + c_2}{b_1}. \end{aligned}$$

In particular,  $T_{20}^1$  is a relative invariant which transforms by a square, so its sign is fixed. The coframing (4.5) is 1-adapted, and if we set

$$dr_0 = r_{01} \omega^1 + r_{02} \omega^2 + r_{03} \omega^3 + r_{00} \phi,$$

it has  $T_{20}^1 = \frac{r + r_{00}}{r}$ . The condition that each fiber of  $\Sigma$  be a strongly convex curve enclosing the origin is exactly the condition that this quantity be positive (see Lemma 7.3 for a proof), so we can assume that  $T_{20}^1 > 0$ . (4.8) then implies that we can adapt coframings to arrange that

$$T_{20}^1 = 1, \quad T_{12}^1 = T_{13}^3 = 0.$$



A coframing satisfying this condition will be called *2-adapted*. Any two 2-adapted coframings on  $\Sigma$  vary by a transformation of the form

$$(4.9) \quad \begin{bmatrix} \tilde{\eta}^1 \\ \tilde{\eta}^2 \\ \tilde{\eta}^3 \\ \tilde{\phi} \end{bmatrix} = \begin{bmatrix} 1 & 0 & a_2 & 0 \\ 0 & \varepsilon & b_2 & 0 \\ 0 & 0 & \varepsilon & 0 \\ -\varepsilon a_2 & -2b_2 & c_3 & \varepsilon \end{bmatrix}^{-1} \begin{bmatrix} \bar{\eta}^1 \\ \bar{\eta}^2 \\ \bar{\eta}^3 \\ \bar{\phi} \end{bmatrix}$$

with  $\varepsilon = \pm 1$ . The set of all 2-adapted coframings forms a principal fiber bundle  $\mathcal{B}_2 \subset \mathcal{B}_1$ , with structure group  $G_2$  consisting of all matrices of the form (4.9). When restricted to  $\mathcal{B}_2$ , the connection forms  $\beta_1, \gamma_1 + \alpha_2, \gamma_2 + 2\beta_2$  become semi-basic. By adding multiples of the semi-basic forms to the connection forms so as to absorb as much of the torsion as possible, we can arrange that the structure equations of  $\mathcal{B}_2$  take the form

$$(4.10) \quad \begin{bmatrix} d\eta^1 \\ d\eta^2 \\ d\eta^3 \\ d\phi \end{bmatrix} = - \begin{bmatrix} 0 & 0 & \alpha_2 & 0 \\ 0 & 0 & \beta_2 & 0 \\ 0 & 0 & 0 & 0 \\ -\alpha_2 & -2\beta_2 & \gamma_3 & 0 \end{bmatrix} \wedge \begin{bmatrix} \eta^1 \\ \eta^2 \\ \eta^3 \\ \phi \end{bmatrix} + \begin{bmatrix} \eta^2 \wedge \phi \\ -\eta^1 \wedge \phi + T_{12}^2 \eta^1 \wedge \eta^2 + T_{20}^2 \eta^2 \wedge \phi \\ \eta^1 \wedge \eta^2 + T_{23}^3 \eta^2 \wedge \eta^3 - T_{12}^2 \eta^3 \wedge \eta^1 + T_{20}^2 \eta^3 \wedge \phi \\ T_{12}^0 \eta^1 \wedge \eta^2 + T_{10}^0 \eta^1 \wedge \phi + T_{20}^0 \eta^2 \wedge \phi \end{bmatrix}.$$

Moreover, we have

$$\begin{aligned} 0 &\equiv d(d\eta^1) \mod \eta^3 \\ &\equiv (T_{12}^2 + T_{10}^0) \eta^1 \wedge \eta^2 \wedge \phi; \end{aligned}$$

therefore,  $T_{10}^0 = -T_{12}^2$ .

Under a transformation of the form (4.9), we have

$$(4.11) \quad \begin{aligned} \tilde{T}_{12}^2 &= T_{12}^2 + \varepsilon(a_2 T_{20}^2 + b_2) \\ \tilde{T}_{23}^3 &= \varepsilon T_{23}^3 + 2b_2 T_{20}^2 - a_2, \end{aligned}$$

so we can adapt coframings to arrange that

$$T_{12}^2 = T_{23}^3 = 0.$$

A coframing satisfying this condition will be called *3-adapted*. Any two 3-adapted coframings on  $\Sigma$  vary by a transformation of the form

$$(4.12) \quad \begin{bmatrix} \tilde{\eta}^1 \\ \tilde{\eta}^2 \\ \tilde{\eta}^3 \\ \tilde{\phi} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \varepsilon & 0 & 0 \\ 0 & 0 & \varepsilon & 0 \\ 0 & 0 & c_3 & \varepsilon \end{bmatrix}^{-1} \begin{bmatrix} \bar{\eta}^1 \\ \bar{\eta}^2 \\ \bar{\eta}^3 \\ \bar{\phi} \end{bmatrix}.$$

The set of all 3-adapted coframings forms a principal fiber bundle  $\mathcal{B}_3 \subset \mathcal{B}_2$ , with structure group  $G_3$  consisting of all matrices of the form (4.12). When restricted to  $\mathcal{B}_3$ , the connection forms  $\alpha_2, \beta_2$  become semi-basic. By adding multiples of the semi-basic forms to the connection forms so as to absorb as much of the torsion as possible, we can arrange that the structure equations of  $\mathcal{B}_3$  take the form

$$(4.13) \quad \begin{bmatrix} d\eta^1 \\ d\eta^2 \\ d\eta^3 \\ d\phi \end{bmatrix} = - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_3 & 0 \end{bmatrix} \wedge \begin{bmatrix} \eta^1 \\ \eta^2 \\ \eta^3 \\ \phi \end{bmatrix} + \begin{bmatrix} \eta^2 \wedge \phi + T_{13}^1 \eta^1 \wedge \eta^3 + T_{23}^1 \eta^2 \wedge \eta^3 + T_{30}^1 \eta^3 \wedge \phi \\ -\eta^1 \wedge \phi + T_{20}^2 \eta^2 \wedge \phi + T_{13}^2 \eta^1 \wedge \eta^3 + T_{23}^2 \eta^2 \wedge \eta^3 + T_{30}^2 \eta^3 \wedge \phi \\ \eta^1 \wedge \eta^2 + T_{20}^2 \eta^3 \wedge \phi \\ T_{12}^0 \eta^1 \wedge \eta^2 - T_{30}^1 \eta^1 \wedge \phi + T_{20}^0 \eta^2 \wedge \phi \end{bmatrix}.$$

(The coefficients  $T_{12}^0, T_{20}^0$  in (4.13) are slightly modified from those in (4.10).) Moreover, we have

$$\begin{aligned} 0 &\equiv d(d\eta^3) \mod \phi \\ &\equiv -(T_{13}^1 + T_{23}^2 + T_{20}^2 T_{12}^0) \eta^1 \wedge \eta^2 \wedge \eta^3; \end{aligned}$$

therefore, we can set

$$T_{13}^1 = -\frac{1}{2} T_{20}^2 T_{12}^0 - A_2, \quad T_{23}^2 = -\frac{1}{2} T_{20}^2 T_{12}^0 + A_2$$

for some function  $A_2$  on  $\mathcal{B}_3$ . (The reason for this choice of notation will shortly become apparent.)

Under a transformation of the form (4.12), we have

$$(4.14) \quad \begin{aligned} \tilde{T}_{23}^1 &= T_{23}^1 + \varepsilon c_3 \\ \tilde{T}_{13}^2 &= T_{13}^2 - \varepsilon c_3 \end{aligned}$$

so we can adapt coframings to arrange that

$$T_{23}^1 = T_{13}^2 = A_1$$

for some function  $A_1$ . A coframing satisfying this condition will be called *4-adapted*. Any two 4-adapted coframings on  $\Sigma$  vary by a transformation of the form

$$(4.15) \quad \begin{bmatrix} \tilde{\eta}^1 \\ \tilde{\eta}^2 \\ \tilde{\eta}^3 \\ \tilde{\phi} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \varepsilon & 0 & 0 \\ 0 & 0 & \varepsilon & 0 \\ 0 & 0 & 0 & \varepsilon \end{bmatrix}^{-1} \begin{bmatrix} \bar{\eta}^1 \\ \bar{\eta}^2 \\ \bar{\eta}^3 \\ \bar{\phi} \end{bmatrix}.$$

The set of all 4-adapted coframings forms a principal fiber bundle  $\mathcal{B}_4 \subset \mathcal{B}_3$ , with structure group  $G_4 = \mathbb{Z}/2\mathbb{Z}$ .  $\mathcal{B}_4$  is thus a double cover of  $\Sigma$ , and the 1-forms  $(\eta^1, \eta^2, \eta^3, \phi)$  form a

canonical coframing on  $\mathcal{B}_4$ . When restricted to  $\mathcal{B}_4$ , the last remaining connection form  $\gamma_3$  becomes semi-basic, and the structure equations of  $\mathcal{B}_4$  take the form

$$(4.16) \quad \begin{bmatrix} d\eta^1 \\ d\eta^2 \\ d\eta^3 \\ d\phi \end{bmatrix} = \begin{bmatrix} \eta^2 \wedge \phi - (A_2 + \frac{1}{2}T_{20}^2 T_{12}^0) \eta^1 \wedge \eta^3 + A_1 \eta^2 \wedge \eta^3 + T_{30}^1 \eta^3 \wedge \phi \\ -\eta^1 \wedge \phi + T_{20}^2 \eta^2 \wedge \phi + A_1 \eta^1 \wedge \eta^3 + (A_2 - \frac{1}{2}T_{20}^2 T_{12}^0) \eta^2 \wedge \eta^3 + T_{30}^2 \eta^3 \wedge \phi \\ \eta^1 \wedge \eta^2 + T_{20}^2 \eta^3 \wedge \phi \\ T_{12}^0 \eta^1 \wedge \eta^2 + T_{13}^0 \eta^1 \wedge \eta^3 + T_{23}^0 \eta^2 \wedge \eta^3 - T_{30}^1 \eta^1 \wedge \phi + T_{20}^0 \eta^2 \wedge \phi + T_{30}^0 \eta^3 \wedge \phi \end{bmatrix}.$$

Finally, we differentiate equations (4.16) in order to find any remaining relations among the torsion functions. Setting

$$dT_{jk}^i = T_{jk,1}^i \eta^1 + T_{jk,2}^i \eta^2 + T_{jk,3}^i \eta^3 + T_{jk,0}^i \phi,$$

and computing  $d(d\eta^3) = 0$  yields

$$(4.17) \quad \begin{aligned} T_{20,1}^2 &= T_{30}^2 + T_{20}^2 T_{30}^1 \\ T_{20,2}^2 &= -(T_{30}^1 + T_{20}^2 T_{20}^0). \end{aligned}$$

Then computing  $d(d\eta^2) \equiv 0 \pmod{\eta^3}$  yields

$$T_{20}^0 = -2T_{30}^2.$$

Further differentiation yields only differential equations for the torsion functions and no new functional relations.

If we rename the  $T_{jk}^i$  as follows:

$$\begin{aligned} T_{20}^2 &= I \\ T_{30}^1 &= J_1 \\ T_{30}^2 &= J_2 \\ T_{12}^0 &= K \\ T_{30}^0 &= S_0 \\ T_{23}^0 &= S_1 \\ T_{13}^0 &= -S_2, \end{aligned}$$

then the structure equations on  $\mathcal{B}_4$  become

$$(4.18) \quad \begin{bmatrix} d\eta^1 \\ d\eta^2 \\ d\eta^3 \\ d\phi \end{bmatrix} = \begin{bmatrix} \eta^2 \wedge \phi + A_1 \eta^2 \wedge \eta^3 + (A_2 + \frac{1}{2}IK) \eta^3 \wedge \eta^1 + J_1 \eta^3 \wedge \phi \\ -\eta^1 \wedge \phi + (A_2 - \frac{1}{2}IK) \eta^2 \wedge \eta^3 - A_1 \eta^3 \wedge \eta^1 + J_2 \eta^3 \wedge \phi + I \eta^2 \wedge \phi \\ \eta^1 \wedge \eta^2 + I \eta^3 \wedge \phi \\ S_0 \eta^3 \wedge \phi + S_1 \eta^2 \wedge \eta^3 + S_2 \eta^3 \wedge \eta^1 - J_1 \eta^1 \wedge \phi - 2J_2 \eta^2 \wedge \phi + K \eta^1 \wedge \eta^2 \end{bmatrix}.$$

(Compare with the sub-Riemannian structure equations (2.1).)

Our first result is that  $I$  is the fundamental invariant that determines whether or not a sub-Finsler structure is sub-Riemannian:

**Theorem 4.1.** *The sub-Finsler structure  $\Sigma$  is the unit circle bundle for a sub-Riemannian metric if and only if  $I \equiv 0$ .*

*Proof.* One direction is trivial: if  $\Sigma = \Sigma_1$  for some sub-Riemannian metric, then the canonical coframing which we have constructed on  $\Sigma$  is simply

$$(\eta^1, \eta^2, \eta^3, \phi) = (\omega^1, \omega^2, \omega^3, \alpha),$$

and so the structure equations (4.18) must reduce to (2.1); therefore,  $I \equiv 0$ .

Now suppose that  $I \equiv 0$ . Then

$$0 = d(d\eta^3) = (J_1 \eta^2 - J_2 \eta^1) \wedge \eta^3 \wedge \phi;$$

therefore,  $J_1 \equiv J_2 \equiv 0$ . Now computing  $d(d\eta^1) \equiv 0 \pmod{\eta^1}$  shows that

$$dA_1 \equiv (S_0 - 2A_2) \phi \pmod{\eta^1, \eta^2, \eta^3},$$

and computing  $d(d\eta^2) \equiv 0 \pmod{\eta^2}$  shows that

$$dA_1 \equiv (-S_0 - 2A_2) \phi \pmod{\eta^1, \eta^2, \eta^3}.$$

Therefore,  $S_0 \equiv 0$ , and the structure equations (4.18) have the form (2.1). This implies that  $\Sigma$  is the unit circle bundle for a sub-Riemannian metric, as desired.  $\square$

## 5. THE GEODESIC EQUATIONS

In this section we consider the problem of finding geodesics of the sub-Finsler structure. Recall that the sub-Finsler length of a horizontal curve  $\gamma : [a, b] \rightarrow \mathcal{X}$  is given by

$$(5.1) \quad \mathcal{L}(\gamma) = \int_a^b F(\gamma'(t)) dt.$$

Finding critical points of this functional amounts to solving a constrained variational problem. However, care must be taken when computing variations among horizontal curves on a non-integrable rank  $s$  distribution  $D$ . Given a horizontal curve  $\gamma$ , one would like to consider “ $D$ -variational vector fields on  $\gamma$  that vanish at the endpoints,” but in general the existence of such vector fields is far from guaranteed. In fact, this can fail spectacularly: for example, when  $D$  is an Engel system on a 4-manifold  $\mathcal{M}$ ,  $\mathcal{M}$  is foliated by horizontal curves that have no such variations [5].

If such a vector field exists along  $\gamma$ , then  $\gamma$  is said to be *regular*, and the methods outlined in [9] suffice to find the first variation. A horizontal curve for which this fails is called *non-regular* (or *abnormal*). In [10] Lucas Hsu established the following criterion for a curve to be non-regular:

**Theorem 5.1.** (*Hsu, [10]*) *Let  $\mathcal{I} \subset T^*\mathcal{X}$  be the annihilator of the rank  $s$  distribution  $D \subset T\mathcal{X}$ , and let  $\Psi$  be the pullback of the canonical symplectic 2-form on  $T^*\mathcal{X}$  to  $\mathcal{I}$ . A horizontal curve  $\gamma : [a, b] \rightarrow \mathcal{X}$  is non-regular if and only if it has a lifting  $\tilde{\gamma} : [a, b] \rightarrow \mathcal{I}$  that does not intersect the zero section and satisfies  $\tilde{\gamma}'(t) \lrcorner \Psi = 0$  for all  $t \in [a, b]$ .*

In the present case,  $D$  is a contact system on a 3-manifold with  $\mathcal{I} = \text{span}\{\eta^3\}$ , and it is easy to see that in this case all horizontal curves must be regular. In what follows we will therefore use the variational methods described in [9]; our argument closely follows that of [11].

Choose an orientation of  $D$ , and consider the set of coframes in  $\mathcal{B}_4$  that preserve this orientation; for simplicity we will continue to use the notation  $\mathcal{B}_4$  for this set. Every horizontal curve  $\gamma : [a, b] \rightarrow \mathcal{X}$  lifts to an integral curve  $\bar{\gamma} : [a, b] \rightarrow \mathcal{B}_4$  of the differential system  $\bar{\mathcal{I}} = \{\eta^2, \eta^3\}$  with  $\eta^1(\bar{\gamma}'(t)) \neq 0$ . This lift corresponds to choosing a 4-adapted coframing along the horizontal curve so that the vector  $e_1$  dual to  $\eta^1$  points in the direction of the

velocity vector of the curve. The sub-Finsler length of  $\gamma$  is then equal to the integral of the monic one-form  $\eta^1$  along the lifted curve  $\bar{\gamma}$ . The problem of finding critical curves of the sub-Finsler length functional among horizontal curves is thus equivalent to finding critical curves of

$$(5.2) \quad \bar{\mathcal{L}}(\bar{\gamma}) = \int_{\bar{\gamma}} \eta^1$$

among integral curves  $\bar{\gamma}$  of  $\bar{\mathcal{I}} = \{\eta^2, \eta^3\}$  on  $\mathcal{B}_4$ .

**Proposition 5.2.** *The critical curves of  $\bar{\mathcal{L}}$  among integral curves of  $\bar{\mathcal{I}}$  on  $\mathcal{B}_4$  are precisely the projections of integral curves, with transversality condition  $\eta^1 \neq 0$ , of the differential system  $\mathcal{J} = \{\eta^2, \eta^3, \phi - \lambda\eta^1, d\lambda - C\eta^1\}$  on  $\mathcal{Y} \cong \mathcal{B}_4 \times \mathbb{R}$ , where  $\lambda$  is the coordinate on the fiber  $\mathbb{R}$  and  $C = \lambda^2 I + \lambda J_1 + A_2 + \frac{1}{2}IK$ .*

*Proof.* Following the algorithm in [9], we define a submanifold  $\mathcal{Z} \subset T^*\mathcal{B}_4$  as follows: for each  $x \in \mathcal{B}_4$ , let  $\mathcal{Z}_x = \eta^1(x) + \text{span}\{\bar{\mathcal{I}}_x\}$  and let

$$(5.3) \quad \mathcal{Z} = \bigcup_{x \in \mathcal{B}_4} \mathcal{Z}_x.$$

Let  $\zeta$  be the pullback to  $\mathcal{Z}$  of the canonical 1-form on  $T^*\mathcal{B}_4$ . By the “self-reproducing” property of  $\zeta$ , we may write

$$(5.4) \quad \zeta = \eta^1 + \lambda_2 \eta^2 + \lambda_3 \eta^3$$

(where we have suppressed the obvious pullbacks in our notation). According to the general theory described in [9], the critical points of the functional

$$(5.5) \quad \tilde{\mathcal{L}}(\tilde{\gamma}) = \int_{\tilde{\gamma}} \zeta$$

among unconstrained curves  $\tilde{\gamma}$  on  $\mathcal{Z}$  project to critical curves of  $\bar{\mathcal{L}}$  among integral curves  $\bar{\gamma}$  of  $\bar{\mathcal{I}}$  on  $\mathcal{B}_4$ ; moreover, a curve  $\tilde{\gamma}$  on  $\mathcal{Z}$  is a critical curve of  $\tilde{\mathcal{L}}$  if and only if  $\tilde{\gamma}'(t) \lrcorner d\zeta|_{\tilde{\gamma}(t)} = 0$ .

A straightforward computation shows that

$$(5.6) \quad \begin{aligned} d\zeta = & \lambda_2 \phi \wedge \eta^1 - (1 + \lambda_2 I) \phi \wedge \eta^2 - (\lambda_3 I + J_1 + \lambda_2 J_2) \phi \wedge \eta^3 + \lambda_3 \eta^1 \wedge \eta^2 \\ & + (A_2 + \frac{1}{2}IK - \lambda_2 A_1) \eta^3 \wedge \eta^1 + (A_1 - \frac{1}{2}\lambda_2 IK + \lambda_2 A_2) \eta^2 \wedge \eta^3 \\ & + d\lambda_2 \wedge \eta^2 + d\lambda_3 \wedge \eta^3. \end{aligned}$$

By contracting  $d\zeta$  with the vector fields dual to the coframing  $\{\eta^1, \eta^2, \eta^3, \phi, d\lambda_2, d\lambda_3\}$  on  $\mathcal{Z}$ , we find that subject to the condition  $\tilde{\gamma}^* \eta^1 \neq 0$ , the requirement that  $\tilde{\gamma}' \lrcorner d\zeta = 0$  is equivalent to the condition that  $\tilde{\gamma}$  is an integral curve of the system

$$\mathcal{J} = \{\eta^2, \eta^3, \phi - \lambda_3 \eta^1, d\lambda_3 - (\lambda_3^2 I + \lambda_3 J_1 + A_2 + \frac{1}{2}IK) \eta^1\}$$

on the submanifold  $\mathcal{Y} \subset \mathcal{Z}$  defined by  $\lambda_2 = 0$ . (Henceforth we will omit the subscript on  $\lambda_3$ .) Curves satisfying this requirement project to critical curves of the functional  $\bar{\mathcal{L}}$  among integral curves of  $\bar{\mathcal{I}}$  on  $\mathcal{B}_4$ , and thus to local minimizers of the sub-Finsler length functional  $\mathcal{L}$  on  $\mathcal{X}$ . Since every horizontal curve on  $\mathcal{X}$  is regular, every local minimizer arises in this way. □

We will call a unit speed horizontal curve  $\gamma : [a, b] \rightarrow \mathcal{M}$  a *sub-Finsler geodesic* if it has a lift to an integral curve of  $\mathcal{J}$  on  $\mathcal{Y}$ . When  $\gamma$  has unit speed, it lifts to an integral curve of  $\mathcal{J}$  if and only if it satisfies the *geodesic equations*

$$(5.7) \quad \eta^1 = ds, \eta^2 = 0, \eta^3 = 0, \phi = \lambda ds, d\lambda = C ds.$$

## 6. THE JACOBI OPERATOR AND THE SECOND VARIATION

This argument is similar to that given in [11] for the sub-Riemannian case; we will describe it in some detail since [11] is unpublished.

Since the geodesic equations are defined on the bundle  $\mathcal{Y} \cong \mathcal{B}_4 \times \mathbb{R}$ , we will work on  $\mathcal{Y}$  and use the coframing  $\{\eta^1, \eta^2, \eta^3, \eta^4, \eta^5\}$ , where

$$\begin{aligned} \eta^4 &= \phi - \lambda \eta^1 \\ \eta^5 &= d\lambda - (\lambda^2 I + \lambda J_1 + A_2 + \tfrac{1}{2} IK) \eta^1. \end{aligned}$$

The structure equations (4.18) imply that this coframing has structure equations

$$(6.1) \quad \begin{aligned} d\eta^1 &= \eta^2 \wedge \eta^4 - \lambda \eta^1 \wedge \eta^2 + A_1 \eta^2 \wedge \eta^3 + J_1 \eta^3 \wedge \eta^4 - (\lambda J_1 + A_2 + \tfrac{1}{2} IK) \eta^1 \wedge \eta^3 \\ d\eta^2 &= -\eta^1 \wedge \eta^4 + J_2 \eta^3 \wedge \eta^4 + I \eta^2 \wedge \eta^4 - \lambda I \eta^1 \wedge \eta^2 \\ &\quad + (A_2 - \tfrac{1}{2} IK) \eta^2 \wedge \eta^3 + (A_1 - \lambda J_2) \eta^1 \wedge \eta^3 \\ d\eta^3 &= \eta^1 \wedge \eta^2 + I \eta^3 \wedge \eta^4 - \lambda I \eta^1 \wedge \eta^3 \\ d\eta^4 &= \eta^1 \wedge \eta^5 - J_1 \eta^1 \wedge \eta^4 + (S_1 - \lambda A_1) \eta^2 \wedge \eta^3 + (S_0 - \lambda J_1) \eta^3 \wedge \eta^4 \\ &\quad - (\lambda + 2J_2) \eta^2 \wedge \eta^4 + (\lambda^2 + 2\lambda J_2 + K) \eta^1 \wedge \eta^2 \\ &\quad + (J_1 \lambda^2 + (A_2 - S_0 + \tfrac{1}{2} IK) \lambda - S_2) \eta^1 \wedge \eta^3 \\ d\eta^5 &= -(\lambda^2 I + \lambda J_1 + A_2 + \tfrac{1}{2} IK) [\eta^2 \wedge \eta^4 - \lambda \eta^1 \wedge \eta^2 + A_1 \eta^2 \wedge \eta^3 + J_1 \eta^3 \wedge \eta^4 \\ &\quad - (\lambda J_1 + A_2 + \tfrac{1}{2} IK) \eta^1 \wedge \eta^3] \\ &\quad + \eta^1 \wedge d(\lambda^2 I + \lambda J_1 + A_2 + \tfrac{1}{2} IK). \end{aligned}$$

As in the previous section, every horizontal curve  $\gamma$  has a canonical lift to an integral curve  $\tilde{\gamma}$  of the system  $\tilde{\mathcal{I}} = \{\eta^2, \eta^3\}$  on  $\mathcal{B}_4$ . This in turn has a canonical lift to an integral curve  $\tilde{\gamma}$  of the system  $\tilde{\mathcal{I}} = \{\eta^2, \eta^3, \eta^4\}$  on  $\mathcal{Y}$ . The length of  $\gamma$  is equal to the integral of  $\eta^1$  along the lifted curve  $\tilde{\gamma}$ , and  $\gamma$  is a geodesic if and only if  $\tilde{\gamma}$  is an integral curve of the system  $\mathcal{J} = \{\eta^2, \eta^3, \eta^4, \eta^5\}$  on  $\mathcal{Y}$ .

Suppose that  $\gamma : [0, \ell] \rightarrow \mathcal{X}$  is a horizontal curve joining points  $p$  and  $q$  in  $\mathcal{X}$ . If  $\gamma_t$  is a fixed-endpoint variation of  $\gamma$  through horizontal curves, then  $\gamma_t$  lifts to a variation  $\tilde{\gamma}_t$  of  $\tilde{\gamma}$  through integral curves of  $\tilde{\mathcal{I}}$ ; this variation does not necessarily fix endpoints, but it satisfies the condition

$$\pi \circ \tilde{\gamma}_t(0) = p, \quad \pi \circ \tilde{\gamma}_t(\ell) = q,$$

where  $\pi : \mathcal{Y} \rightarrow \mathcal{X}$  is the usual base point projection. A variation  $\tilde{\gamma}_t$  satisfying these conditions will be called an *admissible variation* of  $\tilde{\gamma}$ , and its variational vector field  $\frac{\partial \tilde{\gamma}_t}{\partial t}$  at  $t = 0$  will be called an *infinitesimal admissible variation* along  $\tilde{\gamma}$ .

Now suppose that  $\gamma$  is a geodesic. Let  $\gamma_{t,u}$  be 2-parameter fixed-endpoint variation of  $\gamma$  through horizontal curves, and let  $\tilde{\gamma}_{t,u}$  be its lift to  $\mathcal{Y}$ , with infinitesimal admissible variations

$$V(s) = \frac{\partial}{\partial t} \tilde{\gamma}_{t,u}(s) \big|_{t=u=0}, \quad W(s) = \frac{\partial}{\partial u} \tilde{\gamma}_{t,u}(s) \big|_{t=u=0}.$$

Let  $(e_1, \dots, e_5)$  be the framing dual to the coframing  $(\eta^1, \dots, \eta^5)$  on  $\mathcal{Y}$ , and write

$$V(s) = \sum_{i=1}^5 V_i(s) e_i(s), \quad W(s) = \sum_{i=1}^5 W_i(s) e_i(s).$$

The Hessian  $\mathcal{L}_{**}(V, W)$  of the length functional is, by definition,

$$\mathcal{L}_{**}(V, W) = \frac{\partial^2}{\partial t \partial u} \mathcal{L}(\gamma_{t,u}) \big|_{t=u=0}.$$

**Proposition 6.1.** *Let  $\tilde{\gamma} : [0, \ell] \rightarrow \mathcal{Y}$  be a lifted geodesic with 2-parameter admissible variation  $\tilde{\gamma}_{t,u}$ . The Hessian of the length functional evaluated at the infinitesimal admissible variations  $V = \sum V_i e_i$  and  $W = \sum W_i e_i$  is*

$$\mathcal{L}_{**}(V, W) = \int_0^\ell W_3 J(V_3) ds,$$

where  $J$  is a self-adjoint, fourth-order differential operator on the space of smooth functions on  $[0, \ell]$  given by

$$(6.2) \quad J(u) = \ddot{u} + \frac{d}{ds} (P\dot{u}) + Qu,$$

for certain functions  $P, Q$  along  $\tilde{\gamma}$ .

Here the dots over  $u$  represent derivatives with respect to  $s$ , and the precise definitions of  $P$  and  $Q$  will appear in the proof.

*Proof.* For any smooth function  $f$  on  $\mathcal{Y}$ , we will write

$$df = \sum_{i=1}^5 f_{,i} \eta^i.$$

Differentiating (6.1) yields relations among the derivatives of the invariants of the sub-Finsler structure, and these relations must be taken into account in the computations that follow.

As in [11], the Hessian  $\mathcal{L}_{**}(V, W)$  is equal to the integral

$$\int_{\tilde{\gamma}} W \lrcorner d(V \lrcorner d\zeta),$$

where  $\zeta = \eta^1 + \lambda \eta^3$ . A long but straightforward computation shows that along  $\tilde{\gamma}$ , the integrand  $W \lrcorner d(V \lrcorner d\zeta)$  is equivalent modulo  $\mathcal{J}$  to

$$\begin{aligned}
& \left[ W_2 \left( \dot{V}_4 + A_1 \dot{V}_3 - (\lambda^2 + 2J_2\lambda + A_1 + K)V_2 \right. \right. \\
& \quad \left. \left. + \left( -J_1\lambda^2 + (A_1I - \tfrac{1}{2}IK - A_2 + S_0)\lambda + S_2 \right) V_3 + J_1V_4 - V_5 \right) \right. \\
& \quad + W_3 \left( -\dot{V}_5 - A_1 \dot{V}_2 + (I\lambda + J_1)\dot{V}_4 \right. \\
& \quad \left. + \left( -J_1\lambda^2 + (A_1I - \tfrac{1}{2}IK - A_2 + S_0)\lambda + S_2 - A_{1,1} \right) V_2 \right. \\
& \quad \left. + \left( (S_0I + I_{,3} + S_{0,4})\lambda^2 + ((A_1 + K)J_2 + J_1S_0 - S_1 - S_{0,1} + S_{2,4})\lambda \right. \right. \\
(6.3) \quad & \quad \left. \left. + \left( \tfrac{1}{4}I^2K^2 + \tfrac{1}{2}(IK_{,3} + I_{,3}K) + A_1^2 + A_2^2 + A_2IK + J_1S_2 + A_{2,3} \right) \right) V_3 \right. \\
& \quad \left. + (I_{,4}\lambda^2 + (IJ_1 + J_{1,4})\lambda + \tfrac{1}{2}(I^2K + KI_{,4} - IA_{1,4}) \right. \\
& \quad \left. - A_1I^2 - A_2I - 2S_0I + J_1^2 - A_1 + A_{2,4} \right) V_4 \\
& \quad \left. + I\lambda V_5 \right) \\
& \quad + W_4 \left( -\dot{V}_2 - (I\lambda + J_1)\dot{V}_3 + J_1V_2 + (-I^2\lambda^2 - (IJ_1 + J_2)\lambda + A_1)V_3 - V_4 \right) \\
& \quad \left. + W_5 \left( \dot{V}_3 - V_2 + I\lambda V_3 \right) \right] \eta^1.
\end{aligned}$$

(This computation takes into account the fact that along  $\tilde{\gamma}$ ,  $V_{i,1} = \dot{V}_i$ .)

Now define  $\Gamma(t, u, s) = \tilde{\gamma}_{t,u}(s)$ . Since each curve  $\tilde{\gamma}_{t,u}$  is an integral curve of  $\tilde{\mathcal{L}}$ , we have

$$\Gamma^* \begin{bmatrix} \eta^1 \\ \eta^2 \\ \eta^3 \\ \eta^4 \\ \eta^5 \end{bmatrix} = \begin{bmatrix} V_1(t, u, s) dt + W_1(t, u, s) du + Y_1(t, u, s) ds \\ V_2(t, u, s) dt + W_2(t, u, s) du \\ V_3(t, u, s) dt + W_3(t, u, s) du \\ V_4(t, u, s) dt + W_4(t, u, s) du \\ V_5(t, u, s) dt + W_5(t, u, s) du + Y_5(t, u, s) ds \end{bmatrix}$$

for some functions  $V_i, W_i, Y_i$  satisfying (with a minor abuse of notation)  $V_i(0, 0, s) = V_i(s)$ ,  $W_i(0, 0, s) = W_i(s)$ . Since  $\tilde{\gamma}$  is the lift of a geodesic, we also have

$$Y_1(0, 0, s) = 1, \quad Y_5(0, 0, s) = 0, \quad \frac{\partial Y_1}{\partial t}(0, 0, s) = \frac{\partial Y_1}{\partial u}(0, 0, s) = 0.$$

When the structure equations (6.1) are pulled back by  $\Gamma$  and then restricted to  $\tilde{\gamma}$ , they imply that

$$\begin{aligned}
(6.4) \quad & \dot{V}_1 = -\lambda V_2 - (J_1\lambda + A_2 + \tfrac{1}{2}IK)V_3 \\
& \dot{V}_2 = -I\lambda V_2 + (-J_2\lambda + A_1)V_3 - V_4 \\
& \dot{V}_3 = V_2 - I\lambda V_3 \\
& \dot{V}_4 = (\lambda^2 + 2J_2\lambda + K)V_2 + (J_1\lambda^2 + (A_2 + \tfrac{1}{2}IK - S_0)\lambda - S_2)V_3 - J_1V_4 + V_5.
\end{aligned}$$

The third equation in (6.4) implies that  $V_2 = \dot{V}_3 + I\lambda V_3$ . The first equation in (6.4) can then be written as

$$\frac{d}{ds}(V_1 + \lambda V_3) = 0.$$



Since  $V$  is an admissible infinitesimal variation,  $V_1, V_2$ , and  $V_3$  must vanish at the endpoints of  $\tilde{\gamma}$ , and it follows that  $V_1 = -\lambda V_3$ . Equations (6.4) can now be used to express  $V_1, V_2, V_4$ , and  $V_5$  in terms of  $V_3$  and its derivatives, as follows:

$$\begin{aligned}
V_1 &= -\lambda V_3 \\
V_2 &= \dot{V}_3 + I\lambda V_3 \\
V_4 &= -\ddot{V}_3 - 2I\lambda\dot{V}_3 + [-(2I^2 + I_{,4})\lambda^2 - (2IJ_1 + 2J_2)\lambda + (A_1 - IA_2 - \tfrac{1}{2}I^2K)]V_3 \\
(6.5) \quad V_5 &= -\ddot{V}_3 - (2I\lambda + J_1)\ddot{V}_3 \\
&\quad + [-(4I^2 + 3I_{,4} + 1)\lambda^2 - (8IJ_1 + 6J_2)\lambda + (A_1 - 3IA_2 - \tfrac{3}{2}I^2K - K)]\dot{V}_3 \\
&\quad + [-(4I^3 + 6II_{,4} + I)\lambda^3 \\
&\quad - (12I^2J_1 + 5I_{,4}J_1 + I_{,41} + 2IJ_{1,4} + 8IJ_2 + 2J_{2,4} + J_1)\lambda^2 \\
&\quad + (-3IKI_{,4} - 3A_2I_{,4} + 2A_{1,4} + I^2A_{1,4} - 2IA_{2,4} + 4A_1I + 2I^3A_1 \\
&\quad + 3A_2 - 2I^3K - 8IJ_1^2 - 8J_1J_2 - \tfrac{3}{2}IK + 3S_0 + 4I^2S_0)\lambda \\
&\quad + (A_{1,1} - IA_{2,1} - \tfrac{1}{2}I^2K_{,1} + A_1J_1 - 4IA_2J_1 - 3A_2J_2 - \tfrac{5}{2}I^2J_1K \\
&\quad - 2IJ_2K + S_2)]V_3.
\end{aligned}$$

When these expressions are substituted into the integrand (6.3), it takes the form

$$W_3(\ddot{V}_3 + \frac{d}{ds}(P\dot{V}_3) + QV_3),$$

where

$$\begin{aligned}
P &= (2I^2 + 4I_{,4} + 1)\lambda^2 + 8(J_1 I + J_2)\lambda + (2K I^2 + 4A_2 I + K - 3A_1) \\
Q &= [4I^4 + 2(8I_{,4} + 1)I^2 + (5I_{,4}^2 + I_{,4})]\lambda^4 \\
&+ [20J_1 I^3 + (14J_2 + 10J_{1,4})I^2 + (40J_1 I_{,4} + 6I_{,41} + 4J_1 + 8J_{2,4})I + (12J_2 I_{,4} + 6J_{1,4} I_{,4} + J_2 + J_{1,4})]\lambda^3 \\
&+ [(3K - 8A_1)I^4 - (10A_2 + 16S_0 + 4A_{1,4})I^3 + (16K I_{,4} - 10A_1 I_{,4} - 19A_1 + 8A_{2,4} + 40J_1^2 + K)I^2 \\
&\quad + (5A_2 I_{,4} - 4A_{1,4} I_{,4} - 16S_0 I_{,4} + 42J_1 J_2 + 18J_1 J_{1,4} + 2J_{1,41} - 12A_2 - \frac{13}{2}A_{1,4} - 6S_0)I \\
&\quad + (5K I_{,4}^2 + 18J_1^2 I_{,4} - 13A_1 I_{,4} + 6A_{2,4} I_{,4} + 2K I_{,4} + 6J_1 I_{,41} + I_{,411} + I_{,3} - 2A_1 + A_{2,4} \\
&\quad + 3J_1^2 + 8J_2^2 + 8J_2 J_{1,4} + 10J_1 J_{2,4} + 2J_{2,41} + S_{0,4})]\lambda^2 \\
&+ [(12K J_1 - 20A_1 J_1 - 2A_{1,1} + K_{,1})I^3 \\
&\quad + (4K J_{1,4} - 9J_1 A_{1,4} - 2A_{2,1} - A_{1,41} - 14J_1 A_2 - 12J_2 A_1 + 9J_2 K - 36J_1 S_0 - 4S_{0,1})I^2 \\
&\quad + (20K J_1 I_{,4} + 4K_{,1} I_{,4} + 4K I_{,41} + 7A_2 J_{1,4} + 4K J_{2,4} + 16J_1 A_{2,4} - 5J_2 A_{1,4} - 3A_{1,1} + 2A_{2,41} \\
&\quad + \frac{3}{2}K_{,1} - 34A_1 J_1 - 4A_2 J_2 + 4K J_1 + 24J_1^3 - 20J_2 S_0 + 2S_2)I \\
&\quad + (14J_1 A_2 I_{,4} + 4A_{2,1} I_{,4} + 8K J_2 I_{,4} + 5A_2 I_{,41} + 7A_2 J_{2,4} - 8J_1 A_{1,4} + 8J_2 A_{2,4} - 3A_{2,1} - 2A_{1,41} \\
&\quad - 4S_{0,1} + S_{2,4} - 18A_1 J_2 - 13A_2 J_1 + 24J_1^2 J_2 + \frac{5}{2}K J_2 - 8J_1 S_0 - S_1)]\lambda \\
&+ [(\frac{3}{4}K^2 - 3A_1 K)I^4 - (\frac{3}{2}K A_{1,4} + 5A_1 A_2 + 3A_2 K + 6K S_0)I^3 \\
&\quad + (\frac{5}{2}K^2 I_{,4} - \frac{5}{2}A_2 A_{1,4} + 3K A_{2,4} + 3J_1 K_{,1} + \frac{1}{2}K_{,11} - 7A_2^2 + K^2 - \frac{13}{2}A_1 K + 9K J_1^2 - 10A_2 S_0)I^2 \\
&\quad + (6A_2 K I_{,4} + 5A_2 A_{2,4} + 4J_1 A_{2,1} - 3K A_{1,4} + A_{2,11} + 3J_2 K_{,1} + \frac{1}{2}K_{,3} - 11A_1 A_2 + 12A_2 J_1^2 \\
&\quad - 3A_2 K + 11J_1 J_2 K - \frac{7}{2}K S_0)I \\
&\quad + (3A_2^2 I_{,4} + \frac{1}{2}K I_{,3} - 5A_2 A_{1,4} + 4J_2 A_{2,1} + A_{2,3} - A_{1,11} - S_{2,1} + 2A_1^2 - 8A_2^2 + 12A_2 J_1 J_2 \\
&\quad - 6A_2 S_0 + 2J_2^2 K + J_1 S_2)].
\end{aligned}$$

Thus the proposition is proved.  $\square$

We saw in the proof of this proposition that any infinitesimal admissible variation  $V = \Sigma_i V_i e_i$  satisfies

$$\dot{V}_3 = V_2 - I\lambda V_3,$$

and that  $V_1, V_2, V_3$  vanish at the endpoints  $\tilde{\gamma}(0), \tilde{\gamma}(\ell)$  of  $\tilde{\gamma}$ . In particular,  $V_3$  vanishes to first order at 0 and  $\ell$ . Let  $C_0^\infty[0, \ell]$  denote the space of smooth functions on  $[0, \ell]$  that vanish to first order at the endpoints, and note that the Jacobi operator  $J$  is formally self-adjoint on  $C_0^\infty[0, \ell]$ .

Define a quadratic form  $\mathcal{Q}(u)$  by

$$\mathcal{Q}(u) = \mathcal{L}_{**}(u, u).$$

Recall that the *index* of  $\mathcal{Q}$  is the dimension of the largest subspace of  $C_0^\infty[0, \ell]$  on which  $\mathcal{Q}$  is negative definite. Because  $J$  is self-adjoint on  $C_0^\infty[0, \ell]$ , its eigenvalues form a countable subset of the real numbers with  $+\infty$  as the only possible cluster point. It follows that  $J$  has only finitely many negative eigenvalues, and that therefore the index of  $\mathcal{Q}$  is finite.

**Definition 6.2.** A point  $c \in (0, \ell)$  is a conjugate point of  $J$  with multiplicity  $m$  if the space of solutions of the system

$$(6.6) \quad J(u) = 0, \quad u(0) = \dot{u}(0) = 0$$

which vanish to first order at  $c$  has dimension  $m > 0$ . The point  $\gamma(c)$  along a geodesic  $\gamma$  is a conjugate point of  $\gamma$  if  $c$  is a conjugate point of the corresponding Jacobi operator  $J$  along  $\tilde{\gamma}$ .

Note that, since  $J$  is a fourth-order operator, the multiplicity of any conjugate point of  $J$  is either one or two.

**Theorem 6.3.** *The index of  $\mathcal{Q}$  is equal to the number of conjugate points of  $J$ , counted with multiplicity.*

*Proof.* Suppose that the index of  $\mathcal{Q}$  is  $n$ , and let

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n < 0$$

be the negative eigenvalues of  $J$ . For each  $s \in (0, \ell]$ , let

$$\Lambda_1(s) \leq \Lambda_2(s) \leq \cdots$$

denote the eigenvalues of the operator  $J$  on the space  $[0, s]$ . (Note that  $\Lambda_i(\ell) = \lambda_i$  for  $i = 1, \dots, n$ .) It follows from general theory (see, e.g., [12]) that each  $\Lambda_i(s)$  is a strictly decreasing, continuous function on  $(0, \ell]$ , with  $\lim_{s \rightarrow 0^+} \Lambda_i(s) = +\infty$ . Therefore, each function  $\Lambda_i(s)$ ,  $i = 1, \dots, n$  has exactly one root  $c_i$ .

These roots

$$0 < c_1 \leq c_2 \leq \cdots \leq c_n < \ell$$

are precisely the conjugate points of  $J$  between 0 and  $\ell$ . To see this, note that by definition, the condition  $\Lambda_i(c_i) = 0$  implies that there exists a function  $u_i \in C_0^\infty[0, c_i]$  satisfying  $J(u_i) = 0$ . By extending  $u_i$  to a solution of  $J(u) = 0$  on  $C_0^\infty[0, \ell]$ , we get a solution to (6.6) which vanishes to first order at  $c_i$ . Conversely, if  $c \in (0, \ell)$  is a conjugate point, then  $J$  has a zero eigenvalue on  $C_0^\infty[0, c]$ ; therefore,  $\Lambda_j(c) = 0$  for some positive integer  $j$ . Since  $\Lambda_j$  is a strictly decreasing function, it follows that  $\Lambda_j(\ell)$  is equal to one of the negative eigenvalues  $\lambda_j$  of  $J$ , and therefore,  $c = c_j$ .  $\square$

**Corollary 6.4.** *A geodesic no longer minimizes length beyond its first conjugate point.*

*Proof.* Suppose that  $\gamma(c)$  is the first conjugate point of the geodesic  $\gamma$ . Theorem 6.3 implies that for every  $\ell > c$ , the index of  $\mathcal{Q}$  on the space  $C_0^\infty[0, \ell]$  is positive, and so there exists a function  $u \in C_0^\infty[0, \ell]$  for which  $\mathcal{Q}(u) < 0$ . Setting  $V_3 = u$  and defining  $V_1, V_2, V_4, V_5$  according to equations (6.5) defines a direction  $V = \sum_{i=1}^5 V_i e_i$  along which the length functional  $\tilde{\mathcal{L}}$  (and hence  $\mathcal{L}$ ) decreases.  $\square$

## 7. SYMMETRIES AND HOMOGENEOUS EXAMPLES

In this section we examine the symmetries of sub-Finsler structures and describe homogeneous examples.

**Definition 7.1.** *Let  $\Sigma$  be a sub-Finsler structure on  $(\mathcal{X}, D)$ . A symmetry of  $\Sigma$  is a diffeomorphism  $\Phi : \mathcal{X} \rightarrow \mathcal{X}$  which satisfies  $\Phi'(\Sigma) = \Sigma$ . (Note that this implies that  $\Phi'(D) = D$  as well.) A symmetry of the  $\mathbb{Z}/2\mathbb{Z}$ -structure  $\mathcal{B}_4$  is a diffeomorphism  $\Psi : \mathcal{B}_4 \rightarrow \mathcal{B}_4$  with the property that*

$$(\Psi^* \eta^1, \Psi^* \eta^2, \Psi^* \eta^3, \Psi^* \phi) = (\eta^1, \eta^2, \eta^3, \phi).$$

A standard argument shows that the map

$$\Phi \rightarrow \Phi'$$

gives a one-to-one correspondence between orientation-preserving symmetries of  $\Sigma$  and symmetries of  $\mathcal{B}_4$ . By a theorem of Kobayashi [13], it follows that the group of symmetries of  $\Sigma$  can be given the structure of a Lie group of dimension at most four.

There are two possible definitions of homogeneity for a sub-Finsler structure: we could say that  $\Sigma \subset T\mathcal{X}$  is homogeneous if its group of symmetries acts transitively on  $\mathcal{X}$ , or we could require the more restrictive condition that this group act transitively on  $\Sigma$ . Both notions are interesting, and we will consider each of them in the remainder of this section.

**7.1. Symmetry groups of dimension four.** First we consider the case where the group of symmetries of  $\Sigma$  is four-dimensional and acts transitively on  $\mathcal{B}_4$ . Since any symmetry must preserve the canonical coframing  $(\eta^1, \eta^2, \eta^3, \phi)$  on  $\mathcal{B}_4$ , it follows that all the torsion functions must be constants. Conversely, if all the torsion coefficients are constants, then the structure equations of  $\mathcal{B}_4$  define a local Lie group structure on  $\mathcal{B}_4$  for which the canonical coframing is left-invariant; this Lie group then acts transitively on  $\mathcal{B}_4$  in the obvious way.

So, suppose that all the torsion functions in the structure equations (4.18) are constant. Then

$$0 = d(d\eta^3) = [-(IJ_1 + J_2)\eta^1 + (J_1 - 2IJ_2)\eta^2] \wedge \eta^3 \wedge \phi;$$

therefore,  $J_1 = J_2 = 0$ . Next we have

$$0 = d(d\eta^1) \equiv -S_2 \eta^1 \wedge \eta^2 \wedge \eta^3 \pmod{\phi}$$

$$0 = d(d\eta^2) \equiv S_1 \eta^1 \wedge \eta^2 \wedge \eta^3 \pmod{\phi};$$

therefore,  $S_1 = S_2 = 0$ . Then

$$0 = d(d\phi) \equiv (S_0 - IK)\eta^1 \wedge \eta^2 \wedge \phi \pmod{\eta^3};$$

therefore,  $S_0 = IK$ . Now

$$0 = d(d\eta^1) = [(2A_1 + IA_2 + \frac{1}{2}I^2K)\eta^1 + (-2IA_1 + 2A_2 - IK)\eta^2] \wedge \eta^3 \wedge \phi;$$

therefore,

$$A_1 = -\frac{I^2K}{I^2 + 2}, \quad A_2 = -\frac{IK(I^2 - 2)}{2(I^2 + 2)}.$$

Finally, we have

$$0 = d(d\eta^2) = \frac{4IK}{I^2 + 2} \eta^1 \wedge \eta^3 \wedge \phi;$$

therefore,  $IK = 0$ . If  $I = 0$ , then  $\Sigma$  is sub-Riemannian; the homogeneous sub-Riemannian structures are classified in [11]. So suppose that  $I \neq 0$ . Then we have  $K = 0$ , and the structure equations (4.18) reduce to

$$\begin{aligned} d\eta^1 &= \eta^2 \wedge \phi \\ d\eta^2 &= -\eta^1 \wedge \phi + I\eta^2 \wedge \phi \\ d\eta^3 &= \eta^1 \wedge \eta^2 + I\eta^3 \wedge \phi \\ d\phi &= 0. \end{aligned} \tag{7.1}$$

Differentiating (7.1) yields no additional restrictions; thus there exists (at least locally) a 1-parameter family of homogeneous sub-Finsler structures which are not sub-Riemannian.

In fact, these structure equations can be integrated explicitly. First, since  $d\phi = 0$ , there exists a function  $\theta$  on  $\Sigma$  such that

$$\phi = d\theta.$$

Next, note that the system  $S = \{\eta^1, \eta^2\}$  is Frobenius (i.e.,  $dS \equiv 0 \pmod{S}$ ); therefore, there exist functions  $x, y$  on  $\Sigma$ , independent from  $\theta$ , such that

$$S = \{dx, dy\},$$

and functions  $a_{ij}$ ,  $i, j = 1, 2$ , such that

$$\begin{aligned}\eta^1 &= a_{11} dx + a_{12} dy \\ \eta^2 &= a_{21} dx + a_{22} dy.\end{aligned}$$

Now the first two equations of (7.1) imply that the  $a_{ij}$  are functions of  $x, y, \theta$  alone, and that

$$\begin{aligned}\frac{\partial a_{11}}{\partial \theta} &= -a_{21} & \frac{\partial a_{12}}{\partial \theta} &= -a_{22} \\ \frac{\partial a_{21}}{\partial \theta} &= a_{11} - I a_{21} & \frac{\partial a_{22}}{\partial \theta} &= a_{12} - I a_{22}.\end{aligned}$$

In other words, the function pairs  $(a_{11}, a_{21})$  and  $(a_{12}, a_{22})$  are each solutions of the system

$$(7.2) \quad \frac{\partial f}{\partial \theta} = -g$$

$$\frac{\partial g}{\partial \theta} = f - I g$$

for functions  $f(x, y, \theta)$ ,  $g(x, y, \theta)$ . The solution of these differential equations depends on the value of  $I$ .

7.1.1. *Case 1:  $I^2 > 4$ .* The general solution of (7.2) in this case is

$$\begin{aligned}f &= c_1(x, y)e^{r_1\theta} + c_2(x, y)e^{r_2\theta} \\ g &= -c_1(x, y)r_1e^{r_1\theta} - c_2(x, y)r_2e^{r_2\theta},\end{aligned}$$

where

$$r_1, r_2 = \frac{1}{2}(-I \pm \sqrt{I^2 - 4}).$$

By modifying  $x$  and  $y$  if necessary, we can assume that

$$\begin{aligned}a_{11} &= c_1(x, y)e^{r_1\theta} & a_{12} &= c_2(x, y)e^{r_2\theta} \\ a_{21} &= -c_1(x, y)r_1e^{r_1\theta} & a_{22} &= -c_2(x, y)r_2e^{r_2\theta}.\end{aligned}$$

Then the first two equations of (7.1) imply that  $c_1$  is a function of  $x$  alone and  $c_2$  is a function of  $y$  alone.

Now the third equation of (7.1) implies that

$$d\eta^3 \equiv c_1(x)c_2(y)e^{-I\theta}\sqrt{I^2 - 4} dx \wedge dy \pmod{\eta^3}.$$

By the Pfaff theorem, there exists a function  $z$  on  $\Sigma$ , independent from  $x, y$ , and  $\theta$ , such that

$$\eta^3 = c_1(x)c_2(y)e^{-I\theta}\sqrt{I^2 - 4} \left( dz + \frac{1}{2}(x dy - y dx) \right).$$

Finally, the third equation of (7.1) now implies that  $c_1$  and  $c_2$  are in fact constants. Without loss of generality, we may assume that  $c_1 = c_2 = 1$ , and that our coframing has the form

$$\begin{aligned}\eta^1 &= e^{r_1\theta} dx + e^{r_2\theta} dy \\ \eta^2 &= -r_1 e^{r_1\theta} dx - r_2 e^{r_2\theta} dy \\ \eta^3 &= e^{-I\theta} \sqrt{I^2 - 4} \left( dz + \frac{1}{2}(x dy - y dx) \right) \\ \phi &= d\theta.\end{aligned}$$

7.1.2. *Case 2:  $I^2 < 4$ .* The general solution of (7.2) in this case is

$$\begin{aligned}f &= e^{-I\theta/2} [c_1(x, y) \cos(r\theta) + c_2(x, y) \sin(r\theta)] \\ g &= \frac{1}{2} e^{-I\theta/2} [c_1(x, y) (I \cos(r\theta) + r \sin(r\theta)) \\ &\quad + c_2(x, y) (I \sin(r\theta) - r \cos(r\theta))],\end{aligned}$$

where

$$r = \frac{1}{2} \sqrt{4 - I^2}.$$

A similar argument to that given above shows that that we can take our coframing to be

$$\begin{aligned}\eta^1 &= e^{-I\theta/2} [\cos(r\theta) dx + \sin(r\theta) dy] \\ \eta^2 &= \frac{1}{2} e^{-I\theta/2} [(I \cos(r\theta) + r \sin(r\theta)) dx + (I \sin(r\theta) - r \cos(r\theta)) dy] \\ \eta^3 &= -r e^{-I\theta} \left( dz + \frac{1}{2}(x dy - y dx) \right) \\ \phi &= d\theta.\end{aligned}$$

7.1.3. *Case 3:  $I = 2$ .* The general solution of (7.2) in this case is

$$\begin{aligned}f &= e^{-t} [-c_1(x, y) t + c_2(x, y) (1 + t)] \\ g &= e^{-t} [c_1(x, y) (1 - t) + c_2(x, y) t].\end{aligned}$$

A similar argument to that given above shows that that we can take our coframing to be

$$\begin{aligned}\eta^1 &= e^{-\theta} [(1 + \theta) dx - \theta dy] \\ \eta^2 &= e^{-\theta} [\theta dx + (1 - \theta) dy] \\ \eta^3 &= e^{-2\theta} \left( dz + \frac{1}{2}(x dy - y dx) \right) \\ \phi &= d\theta.\end{aligned}$$

7.1.4. *Case 4:  $I = -2$ .* The general solution of (7.2) in this case is

$$\begin{aligned}f &= e^t [-c_1(x, y) t + c_2(x, y) (1 - t)] \\ g &= e^t [c_1(x, y) (1 + t) + c_2(x, y) t].\end{aligned}$$

A similar argument to that given above shows that that we can take our coframing to be

$$\begin{aligned}\eta^1 &= e^\theta [(1 - \theta) dx - \theta dy] \\ \eta^2 &= e^\theta [\theta dx + (1 + \theta) dy] \\ \eta^3 &= e^{2\theta} \left( dz + \frac{1}{2}(x dy - y dx) \right) \\ \phi &= d\theta.\end{aligned}$$

Note that none of these four coframings have coordinate expressions which are periodic in the  $\theta$  variable. Consequently, in all four cases the indicatrix for the sub-Finsler metric in each

tangent space fails to be a closed curve. (In fact, these indicatrices are not even connected.) Therefore, these sub-Finsler structures exist “micro-locally” – that is, in a neighborhood of each point in  $\Sigma$  – but not locally. In other words, there is no open set  $\mathcal{U} \subset \mathcal{X}$  for which the sub-Finsler metric is defined on all of  $T\mathcal{U}$ . This is consistent with a theorem of Rund (see [1]) which implies that for any Minkowski norm on a plane  $D_x$ , the average value of  $I$  over the indicatrix must be zero; therefore, if  $I$  is any nonzero constant, the indicatrix cannot possibly be a closed, strongly convex curve in  $D_x$ .

**7.2. Symmetry groups of dimension three.** Now we consider the more inclusive case where the group  $G$  of symmetries of  $\Sigma$  is three-dimensional and acts transitively on  $\mathcal{X}$ . Since  $\Sigma$  is invariant under the action of  $G$ , it is completely determined by the fiber  $\Sigma_x$  at any point  $x \in \mathcal{X}$ . Conversely, if we fix a point  $x \in \mathcal{X}$  and choose any smooth curve  $\Gamma \subset D_x$  which is strongly convex and encloses the origin, then there exists a unique sub-Finsler structure  $\Sigma$  on  $(\mathcal{X}, D)$  which is invariant under the action of  $G$  and satisfies  $\Sigma_x = \Gamma$ .

Thus we conclude that the sub-Finsler structures of this type are generated by choosing a three-dimensional Lie group  $G$  (or a quotient thereof by a discrete subgroup), a 2-plane  $D \subset T_e G$  which is not a Lie subalgebra (so that it is bracket-generating), and a smooth curve  $\Gamma$  in  $D$  which is strongly convex and surrounds the origin.

**Example 7.2.** Let  $\mathcal{H}$  be the Heisenberg group, defined by

$$\mathcal{H} = \left\{ \begin{bmatrix} 1 & y & z + \frac{1}{2}xy \\ 0 & 1 & x \\ 0 & 0 & 1 \end{bmatrix} : x, y, z \in \mathbb{R} \right\} \cong \mathbb{R}^3,$$

and let the contact structure on  $\mathcal{H}$  be the rank two distribution

$$(7.3) \quad D = \left\{ dz + \frac{1}{2}(x dy - y dx) \right\}^\perp.$$

The existence of this global coordinate system makes it easy to describe sub-Finsler geodesics within the Heisenberg group. Moreover, this example is prototypical: by a theorem of Pfaff (see [4]), any contact 3-manifold has local coordinates  $(x, y, z)$  for which the contact system is given by the symmetric normal form (7.3) above.

We can define a homogeneous, flat sub-Riemannian metric on  $(\mathcal{H}, D)$  by declaring the vectors

$$V_1 = \frac{\partial}{\partial x} + \frac{y}{2} \frac{\partial}{\partial z}, \quad V_2 = \frac{\partial}{\partial y} - \frac{x}{2} \frac{\partial}{\partial z}$$

to be orthonormal. Let  $\Sigma_1$  be the unit circle bundle for this sub-Riemannian structure on  $\mathcal{H}$ , and define a coordinate  $\theta$  on  $\Sigma_1$  by the condition that, for  $u \in \Sigma_1$ ,

$$u = (\cos \theta)V_1 + (\sin \theta)V_2.$$

It is straightforward to check that  $V_1, V_2$  are left-invariant, horizontal vector fields on  $(\mathcal{H}, D)$ , and that therefore any scaling function  $r(\theta)$  which depends on  $\theta$  alone defines a homogeneous sub-Finsler structure on  $\mathcal{H}$ . It is also straightforward to check that the coframing

$$\begin{aligned}\bar{\eta}^1 &= \rho^* ((r \cos \theta - r' \sin \theta) dx - (r \sin \theta + r' \cos \theta) dy) \\ \bar{\eta}^2 &= \rho^* \left( \sqrt{r(r+r'')} [(\sin \theta) dx + (\cos \theta) dy] \right) \\ \bar{\eta}^3 &= \rho^* \left( r^{3/2} \sqrt{r+r''} [dz + \tfrac{1}{2}(x dy - y dx)] \right) \\ \bar{\phi} &= \rho^* \left( \frac{\sqrt{r+r''}}{\sqrt{r}} d\theta \right)\end{aligned}$$

on the sub-Finsler structure  $\Sigma$  defined by  $r(\theta)$  is 4-adapted. The invariants for this coframing are

$$\begin{aligned}I &= -\frac{1}{2} \frac{(rr''' + 3r'r'' + 4rr')}{\sqrt{r}(r+r'')^{3/2}}, \\ K &= A_1 = A_2 = J_1 = J_2 = S_0 = S_1 = S_2 = 0.\end{aligned}$$

In this case, the geodesic equations (5.7) can be written as

$$\begin{aligned}(7.4) \quad dx &= \frac{\cos \theta(s)}{r(\theta(s))} ds \\ dy &= -\frac{\sin \theta(s)}{r(\theta(s))} ds \\ dz &= \frac{x(s) \sin \theta(s) + y(s) \cos \theta(s)}{2r(\theta(s))} ds \\ d\theta &= \frac{\sqrt{r(\theta(s))}}{\sqrt{r(\theta(s)) + r''(\theta(s))}} \lambda(s) ds \\ d\lambda &= I \lambda^2(s) ds.\end{aligned}$$

Since the Lie algebra of  $\mathcal{H}$  is solvable, it is no surprise to find that these equations can be solved by quadrature. First we prove a straightforward but useful lemma.

**Lemma 7.3.** *The expression  $r(r+r'')$  is positive and bounded away from zero.*

*Proof.* Let  $x$  be any point in  $\mathcal{H}$ , and let  $\Sigma_x \subset D_x$  be the indicatrix in  $D_x$  corresponding to a sub-Finsler metric  $F$ . The strong convexity of  $F$  is equivalent to the condition that, for any parametrization  $(u(t), v(t))$  of  $\Sigma_x$ ,

$$(7.5) \quad \frac{u''v' - u'v''}{u'v - uv'} > 0$$

everywhere on  $\Sigma_x$  (see [1] for a proof). In particular, for the parametrization

$$u(\theta) = R(\theta) \cos \theta, \quad v(\theta) = R(\theta) \sin \theta,$$

this is equivalent to saying that

$$(7.6) \quad \frac{r+r''}{r} = \frac{R^2 + 2(R')^2 - RR''}{R^2} > 0$$



everywhere on  $\Sigma_x$ , and hence  $r(r + r'') > 0$  as well. (Recall that  $r$  is the *reciprocal* of the radial position function  $R$  defining  $\Sigma$ .) Since  $\Sigma_x$  is compact, this quantity is bounded away from zero. By the homogeneity of  $\mathcal{H}$ , this bound is the same at every point  $x \in \mathcal{H}$ .  $\square$

This lemma and the geodesic equations (7.4) imply the following result.

**Theorem 7.4.** *For any homogeneous sub-Finsler metric  $F$  on the Heisenberg group  $\mathcal{H}$ , the sub-Finsler geodesics are straight lines parallel to the  $xy$ -plane or liftings of simple closed curves in the  $xy$ -plane. In the latter case, the simple closed curves are the curves of minimal Finsler arc length enclosing a given Euclidean area in the plane.*

*Proof.* A curve  $\gamma : [a, b] \rightarrow \mathcal{H}$  is a geodesic if and only if it is an integral curve of the system (7.4). The equations for  $d\theta$  and  $d\lambda$  in (7.4) allow us to write

$$\begin{aligned} \frac{d\lambda}{\lambda} &= I \sqrt{\frac{r + r''}{r}} d\theta \\ &= -\frac{1}{2} \frac{rr''' + 3r'r'' + 4rr'}{r(r + r'')} d\theta \\ &= -\frac{1}{2} \frac{d(r(r + r''))}{r(r + r'')} - \frac{dr}{r}, \end{aligned}$$

and so

$$\lambda = \frac{c\lambda_0}{r\sqrt{r(r + r'')}},$$

where  $\lambda_0, c$  are constants, with  $c = r(0)\sqrt{r(0)(r(0) + r''(0))} > 0$ .

Integral curves of (7.4) corresponding to  $\lambda_0 = 0$  are straight lines parallel to the  $xy$ -plane. If  $\gamma$  is an integral curve corresponding to some  $\lambda_0 \neq 0$ , then we have

$$d\theta = \frac{c\lambda_0}{r(r + r'')} ds.$$

By the preceding lemma, the quantity in the denominator is positive and bounded away from zero; thus  $\theta$  varies monotonically with  $s$ , without bound. We may therefore reparametrize the equations for  $dx, dy$  and  $dz$  in terms of  $\theta$ . If  $(x_0, y_0) = (x(\theta_0), y(\theta_0))$  is any point on the projection of  $\gamma$  to the  $xy$ -plane, then for any other value  $\theta$ , we have

$$\begin{aligned} (7.7) \quad x(\theta) - x_0 &= \frac{1}{c\lambda_0} \int_{\theta_0}^{\theta} (r + r'') \cos t \, dt \\ y(\theta) - y_0 &= -\frac{1}{c\lambda_0} \int_{\theta_0}^{\theta} (r + r'') \sin t \, dt. \end{aligned}$$

Integrating by parts twice shows that

$$\begin{aligned} x(\theta) - x_0 &= \frac{-1}{c\lambda_0} (r(\theta)^2 u'(\theta) - r(\theta_0)^2 u'(\theta_0)) \\ y(\theta) - y_0 &= \frac{1}{c\lambda_0} (r(\theta)^2 v'(\theta) - r(\theta_0)^2 v'(\theta_0)), \end{aligned}$$

where

$$u(\theta) = R(\theta) \cos \theta, \quad v(\theta) = R(\theta) \sin \theta$$

is the parametrization of the indicatrix used in Lemma 7.3. Thus  $x(\theta) - x_0$  and  $y(\theta) - y_0$  are simultaneously zero if and only if

$$\begin{bmatrix} u'(\theta) \\ v'(\theta) \end{bmatrix} = \frac{r(\theta_0)^2}{r(\theta)^2} \begin{bmatrix} u'(\theta_0) \\ v'(\theta_0) \end{bmatrix},$$

and because the indicatrix is strongly convex, this occurs precisely when  $\theta = \theta_0 + 2n\pi$  for any integer  $n$ . Since  $\theta$  is unbounded as a function of arc length, it attains these values. Therefore, the projection of  $\gamma$  onto the  $xy$ -plane is a simple closed curve.

Finally, since  $dz = -\frac{1}{2}(x dy - y dx)$ , Stokes' theorem implies that the difference  $z(\theta) - z_0$  along any horizontal curve in  $\mathcal{H}$  is proportional to the signed area enclosed by the projection of the curve onto the  $xy$ -plane and the line segment connecting  $(x(\theta), y(\theta))$  to  $(x_0, y_0)$ . Thus  $z$  varies monotonically with increasing  $\theta$ , and the projection of  $\gamma$  onto the  $xy$ -plane is the curve of shortest Finsler arc length enclosing a given Euclidean area in the plane.  $\square$

By the last observation in the proof above, the projections of sub-Finsler geodesics are precisely the solutions to the dual of the classical isoperimetric problem known as *Dido's problem*, named after Queen Dido in Virgil's *Aeneid* (see [14]). The classical solutions (using Riemannian arc length) are circles. In the Finsler case, the solution curves need not be circles, as one of the examples below illustrates.

**7.2.1. Randers metrics on  $\mathcal{H}$  and their geodesics.** Consider the Randers-type, homogeneous sub-Finsler metric on  $D$  obtained by choosing the function  $r(\theta)$  to be

$$r(\theta) = 1 + B \cos \theta, \quad 0 < B < 1.$$

The indicatrix  $\Sigma_x$  for this metric in each plane  $D_x$  is the off-center ellipse with polar equation

$$R = \frac{1}{1 + B \cos \theta}.$$

For this metric,

$$I = \frac{3B \sin \theta}{2\sqrt{1 + B \cos \theta}}.$$

The geodesic equations (7.4) can be integrated explicitly in terms of  $\theta$ . When  $\lambda_0 = 0$ , the geodesics are straight lines; when  $\lambda_0 \neq 0$ , integrating yields:

$$\begin{aligned} x(\theta) - x_0 &= \frac{1}{k}(\sin \theta - \sin \theta_0) \\ (7.8) \quad y(\theta) - y_0 &= \frac{1}{k}(\cos \theta - \cos \theta_0) \end{aligned}$$

$$z(\theta) - z_0 = \frac{1}{2k^2}[\theta - \theta_0 - \sin(\theta - \theta_0)] + \frac{1}{2k}[y_0(\sin \theta - \sin \theta_0) - x_0(\cos \theta - \cos \theta_0)],$$

where  $k = \lambda_0 \sqrt{(1 + B \cos \theta_0)^3}$ . These geodesics are liftings of circles of radius  $1/k$  in the  $xy$ -plane. In the limiting case  $B = 0$ , we recover the geodesics of the flat sub-Riemannian metric. When  $0 < B < 1$ , the anisotropy of the indicatrix is manifested in the way that the area of the projected circle in the  $xy$ -plane (and, therefore,  $dz/d\theta$ ) varies with the initial value  $\theta_0$ , unlike in the sub-Riemannian case.

Figure 1 shows some typical geodesics for  $B = \frac{1}{2}$ , starting from  $(x_0, y_0, z_0) = (0, 0, 0)$ , with initial values  $\lambda_0 = 0.3$  and  $\theta_0 = 0, \pi/2$ , and  $\pi$ . For comparison, Figure 2 shows geodesics for the flat sub-Riemannian metric with these same initial values.

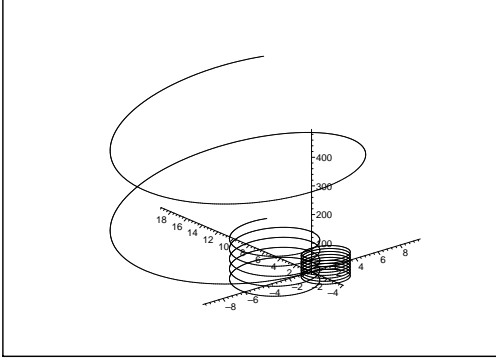


Figure 1: Geodesics of the Randers metric

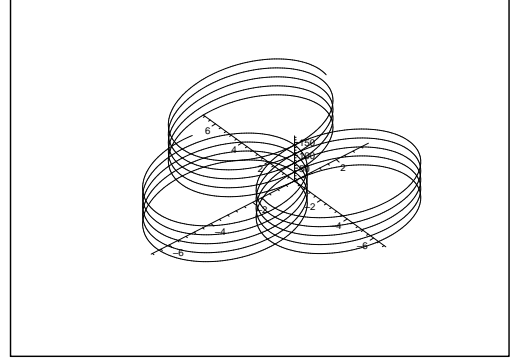


Figure 2: Geodesics of the sub-Riemannian metric

7.2.2. A “limaçon metric” on  $\mathcal{H}$  and its geodesics. For an example in which the geodesics are not liftings of conic sections, consider the sub-Finsler metric whose indicatrix is the convex limaçon with polar equation  $R = 3 + \cos \theta$ , so that

$$r(\theta) = \frac{1}{3 + \cos \theta}.$$

For this metric,

$$I = \frac{-3 \sin \theta (15 \cos \theta + 13)}{2 \sqrt{(9 \cos \theta + 11)^3}}.$$

As always, the geodesics are straight lines when  $\lambda_0 = 0$ , but otherwise they are liftings of curves in the  $xy$ -plane defined by the equations

$$x(\theta) - x_0 = \frac{1}{2L} \left( \frac{\sin \theta (4 \cos \theta + 6)}{(3 + \cos \theta)^2} - \frac{\sin \theta_0 (4 \cos \theta_0 + 6)}{(3 + \cos \theta_0)^2} \right)$$

$$y(\theta) - y_0 = -\frac{1}{L} \left( \frac{9 \cos \theta + 19}{(3 + \cos \theta)^2} - \frac{9 \cos \theta_0 + 19}{(3 + \cos \theta_0)^2} \right),$$

where

$$L = \lambda_0 \frac{\sqrt{9 \cos \theta_0 + 11}}{(3 + \cos \theta_0)^3}, \quad \lambda_0 \neq 0.$$

These curves in the  $xy$ -plane are not circles, nor are they ellipses (or limaçons). Figure 3 shows geodesics for this metric starting from  $(x_0, y_0, z_0) = (0, 0, 0)$ , with initial values  $\lambda_0 = 1$  and  $\theta_0 = 0, \pi/2$ , and  $\pi$ . Figure 4 shows the projections of these curves onto the  $xy$ -plane.

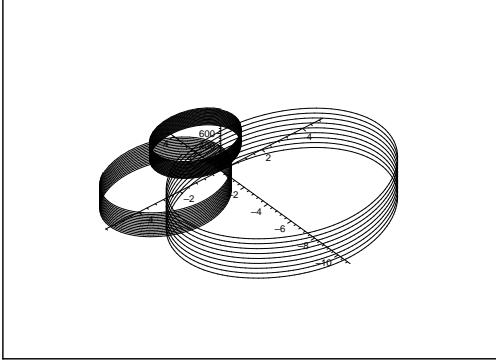


Figure 3: Geodesics of the limaçon metric

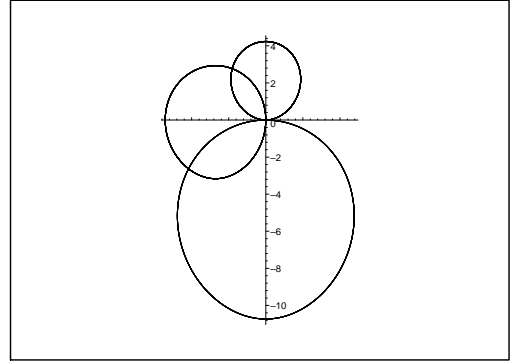


Figure 4: Projections of limaçon metric geodesics onto the  $xy$ -plane

## 8. CONCLUSION

We have only begun to explore sub-Finsler geometry in this paper, and we have every reason to believe that it will become a useful extension of sub-Riemannian geometry, particularly in the context of control theory. In future papers, we plan to investigate higher-dimensional cases (including the important phenomenon of abnormal geodesics), singularities, and other topics likely to be of interest for control theory applications.

## REFERENCES

- [1] D. Bao, S.-S. Chern, and Z. Shen, *An Introduction to Riemann-Finsler Geometry*, Graduate Texts in Mathematics 200, Springer-Verlag, New York, 2000.
- [2] R. Bryant, Finsler surfaces with prescribed curvature conditions, Aisenstadt Lectures, in preparation.
- [3] R. Bryant, Finsler structures on the 2-sphere satisfying  $K = 1$ , *Finsler Geometry (Seattle, WA, 1995)*, 27-41, Contemp. Math., 196, Amer. Math. Soc., Providence, RI, 1996.
- [4] R. Bryant, S. Chern, R. Gardner, H. Goldschmidt, and P. Griffiths, *Exterior Differential Systems*, Math. Sci. Res. Inst. Publ. 18, Springer-Verlag, New York, 1991.
- [5] R. Bryant, L. Hsu, Rigidity of integral curves of rank 2 distributions, *Invent. Math.* 114, 435 - 461 (1993).
- [6] É. Cartan, *Les Espaces Finsler*, Exposés de Géométrie, t. 79, Hermann, Paris, 1934.
- [7] W.-L. Chow, Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung, *Math. Ann.* 117 (1939) 98-105.
- [8] R. Gardner, *The Method of Equivalence and Its Applications*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 58, SIAM, Philadelphia, 1989.
- [9] P. A. Griffiths, *Exterior Differential Systems and the Calculus of Variations*, Birkhäuser, Boston, 1983.
- [10] L. Hsu, *Calculus of variations via the Griffiths formalism*, J. Differential Geom., **36**, 551-589, 1992.
- [11] K. Huguen, *The Geometry of Sub-Riemannian Three-Manifolds*, Ph.D. thesis, Duke University, 1995.
- [12] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1966.
- [13] S. Kobayashi, Theory of connections, *Ann. Mat. Pura Appl.* 43 (1957) 119-194.
- [14] R. Montgomery, *A Tour of Subriemannian Geometries, Their Geodesics and Applications*, Mathematical Surveys and Monographs 91, Amer. Math. Soc., Providence, RI, 2002.
- [15] C. Moseley, *The Geometry of Sub-Riemannian Engel Manifolds*, Ph.D. thesis, University of North Carolina at Chapel Hill, 2001.

DEPARTMENT OF MATHEMATICS, 395 UCB, UNIVERSITY OF COLORADO, BOULDER, CO 80309-0395  
*E-mail address:* `Jeanne.Clelland@colorado.edu`

DEPARTMENT OF MATHEMATICAL SCIENCES, U.S. MILITARY ACADEMY, WEST POINT, NY 10996  
*E-mail address:* `Chris.Moseley@usma.edu`